THE ROLE OF NON-COGNITIVE SKILLS IN EDUCATIONAL PRODUCTION IN SOUTH AFRICA

Heleen Hofmeyr

Dissertation presented for the degree of Doctor of Philosophy in the Faculty Economic and Management Sciences at Stellenbosch University.

Supervisor: Dr Gabrielle Wills

Co-supervisor: Professor Servaas van der Berg

March 2021

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

This dissertation includes three unpublished publications. The development and writing of the papers were the principal responsibility of myself.

ABSTRACT

The role of non-cognitive skills in educational production in low and middle-income countries (LMICs) has largely been overlooked in the international economics of education literature. This constitutes a noteworthy gap in our knowledge of how learning outcomes are produced in LMIC contexts, given the centrality of non-cognitive skills to current education research and policy debates in high-income countries (HICs). This thesis aims to address this gap by investigating the association between non-cognitive skills and learning outcomes in South Africa.

The thesis begins by contextualising the study of non-cognitive skills in the economics of education as a discipline. A case is made for studying the association between non-cognitive skills and learning outcomes in South Africa. In Chapter 2, South African data from the Progress in International Reading Literacy Study (PIRLS) and the Trends in Mathematics and Science Study (TIMSS) were analysed through the lens of "academic resilience" to explore why some students perform above expectations. In accordance with findings from the international literature, a strong association is found between non-cognitive skills and the probability of exceptional performance.

Chapter 3 makes use of reading achievement data from a local study titled Leadership for Literacy to explore potential interaction effects between school functionality and the non-cognitive skill of "grit". The econometric analysis points to evidence of variation in the association between grit and reading achievement by school functionality, with a stronger association estimated for learners in more functional schools. These results provide empirical evidence in support of the hypothesis that school characteristics interact with non-cognitive skills to produce learning outcomes, a relationship that has received scant attention in the international literature on non-cognitive skills in educational production to date.

A natural extension of the results from Chapters 2 and 3 is to explore whether South Africa's gendered educational outcomes can be linked to gender differences in non-cognitive skills. This analysis is undertaken in Chapter 4. Again, the PIRLS and TIMSS data is utilised to model student achievement. Oaxaca-Blinder decomposition analysis is used to investigate how much of South Africa's pro-girl achievement gap in these datasets can be explained by gender differences in observable characteristics, with a particular focus on the contribution of gender differences in non-cognitive skills. Overall, the analysis in this chapter illustrates clearly how focusing on non-cognitive skills as predictors of learning outcomes can enhance our understanding of hitherto unexplained features of South Africa's educational performance, such as the country's large and persistent pro-girl achievement gap.

Together, these results suggest non-cognitive skills are an important input in the educational production process, even in contexts of severe resource deprivation which characterise a large part of the South African education system. This evidence makes an important contribution to local education policy and

practice, as it suggests that targeting non-cognitive skills may be a powerful but hitherto unexplored policy lever for raising learning outcomes in the country.

OPSOMMING

Die internasionale ekonomie van onderwys het tot dusver min aandag gegee aan die rol van nie-kognitiewe vaardighede in die bereiking van leeruitkomstes in lae- en middelinkomste lande. Dit vorm ‘n opvallende leemte in ons kennis van hoe leeruitkomstes bepaal word in lae-en middelinkomste lande, gegewe die sentraliteit van nie-kognitiewe vaardighede tot huidige navorsings- en beleidsdebatte in hoëinkomste lande. Hierdie proefskrif beoog om daardie leemte aan te spreek deur die verhouding tussen nie-kognitiewe vaardighede en leeruitkomstes in Suid-Afrika te ondersoek.

Die proefskrif begin deur die studie van nie-kognitiewe vaardighede binne die ekonomie van onderwys te kontekstualiseer. ‘n Motivering word verskaf vir die belangrikheid van die bestudering van hierdie vaardighede binne die Suid-Afrikaanse konteks. Data van PIRLS en TIMSS word in Hoofstuk 2 gebruik om te bepaal watter faktore verband hou met uitsonderlike prestasie in geletterheid en wiskunde. Daar word bevind - in ooreenstemming met internasionale literatuur - dat nie-kognitiewe vaardighede sterk verband hou met die waarskynlikheid dat ‘n leerder uitsonderlik presteer.

Hoofstuk 3 maak gebruik van leesprestasiedata uit ‘n plaaslike studie genaamd “Leadership for Literacy” om moontlike interaksie-effekte tussen skoolfunksionaliteit en deursettingsvermoë te bestudeer. Die ekonometriese analise wys na variasie in die verband tussen deursettingsvermoë en leesprestasie in skole met verskillende vlakke van funksionaliteit, met ‘n sterker verband by leerders in meer funksionele skole. Hierdie resultate verskaf empiriese ondersteuning vir die hipotese dat skooleienskappe en nie-kognitiewe vaardighede op mekaar inwerk om leeruitkomstes te bepaal, ‘n verband wat tans min aandag geniet in die internasionale literatuur oor nie-kognitiewe vaardighede in die produksie van leeruitkomstes.

‘n Natuurlike uitbreiding van die resultate van Hoofstukke 2 en 3 is om te ondersoek tot watter mate geslagsverskille in nie-kognitiewe vaardighede bydra tot geslagsverskille in leeruitkomstes in Suid-Afrika. Hoofstuk 4 handel oor hierdie kwessie. PIRLS en TIMSS data word weer gebruik om leerderprestasie te modelleer. Die Oaxaca-Blinder ontledingsanalise word gebruik om te bepaal watter proporsie van dogters se voordeel in terme van skoolprestasie verduidelik kan word met verwysing na geslagsverskille in waarneembare eienskappe, met ‘n spesifieke fokus op die bydrae van geslagsverskille in nie-kognitiewe vaardighede. Die analise in hierdie hoofstuk dui daarop dat ‘n fokus op nie-kognitiewe vaardighede ons kan help om sekere kenmerke van die Suid-Afrikaanse leeruitkomslandskap beter te verstaan – soos waarom dogters beter doen as seuns.

Die gevolgtrekking is dat nie-kognitiewe vaardighede na alle waarskynlikheid ‘n belangrike faktor is in die bereiking van leeruitkomstes, selfs in kontekste van ernstige hulpbronontneming wat ‘n groot deel van Suid-Afrika se onderwysstelsel karaktiseer. Hierdie gevolgtrekking maak ‘n belangrike bydrae

tot plaaslike onderwysbeleid en -praktyk, aangesien dit suggereer dat die teiken van nie-kognitiewe vaardighede ‘n kragtige maar onontdekte beleidsopsie kan wees om leeruitkomstes in die land te verbeter.

ACKNOWLEDGEMENTS

The completion of this thesis would not have been possible without the input and support of my supervisors, the broader Research on Socio-Economic Policy (ReSEP) group, or, of course, my friends and family.

I first met Professor Servaas van der Berg as an undergraduate student, and knew from the very first introductory lecture in the economics of education that I wanted to devote my life to research in education and socio-economic policy. That conviction, and the support and guidance that Professor Van der Berg would later provide as supervisor of my postgraduate studies, carried me through the innumerable fears and anxieties of my postgraduate journey. Prof Van der Berg, thank you for your steadfast belief in me. Thank you for encouraging me to pursue a PhD, and for providing the financial support that allowed me to do so. Your intellectual guidance and support have carried me through my worst moments of self-doubt.

Dr Gabrielle Wills, it has been a joy to work with you during the past three years. You challenged me to do better than I ever thought I could, without making me feel inadequate or incapable. Thank you for taking seriously my “out there” idea of looking at non-cognitive skills in education. Your constant feedback and support have kept any potential feelings of isolation that usually accompany a PhD at bay. I am so lucky to have had the privilege of being supervised by you.

To my parents, Murray and Lientjie Hofmeyr: thank you for providing the firm foundation beneath my feet. Thank you for always encouraging me to think critically, and to use my privilege to contribute to a more just society. Thank you to my grandmother, Joan Hofmeyr, who has been my cheerleader throughout the PhD process.

Adaiah Lilenstein, thank you for welcoming me so wholeheartedly into your life from day one. I cannot imagine doing a PhD without you. Thank you for sharing in so many important moments in my life these past three years – the rants, jokes, uncertainty, anxiety, heartbreaks, tears, and adventures. You have been a source of constant support and affirmation, and I am forever grateful to you.

DEDICATION

I dedicate this work to Andri Burnett, Vasti Calitz, and Luami Calitz. You have loved me and supported me in the best way I have ever known. The home you have created has made it possible for me to dedicate myself wholeheartedly to my PhD, knowing that all my other needs were taken care of. Your enthusiasm, humour, ambition, and dedication inspires me to be the best version of myself, and to get the most out of life. Thank you for carrying me through this PhD, and everything else.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

The role of non-cognitive skills in educational production in low and middle-income countries (LMICs) has largely been overlooked in the international economics of education literature. This constitutes a noteworthy gap in our knowledge of how learning outcomes are produced in LMIC contexts, given the centrality of non-cognitive skills to current education research and policy debates in high-income countries (HICs). This thesis attempts to address this gap by investigating the relationship between non-cognitive skills and learning outcomes in South Africa. The results from all three chapters suggest non-cognitive skills are an important input in the educational production process, even in contexts of severe resource deprivation which characterise a large part of the education system in South Africa. This evidence makes an important contribution to the international literature on the role of non-cognitive skills in education, as it suggests that non-cognitive skills may be important predictors of learning outcomes in LMICs, something that has largely been overlooked in the international economics of education literature.

1.1. BACKGROUND AND MOTIVATION

> "The multiple nature of skills is often ignored in many public policy discussions. For example, policy discussions surrounding education and the output of schools often focus on measuring, enhancing, and rewarding cognitive ability measured using achievement tests… An important lesson from the recent economics of human development is that cognitive skills are only part of what is required for success in life. Personality skills - that is, 'soft skills', such as trust, altruism, reciprocity, perseverance, attention, motivation, self-confidence, and personal health - are also important. Health and mental health are essential skills; so too are the abilities to make wise decisions, to guide one's life by reflective reason, and to plan ahead. These skills are often neglected in scientific analyses and policy discussions alike" (Heckman and Corbin, 2016: 345).

This perspective of Heckman and Corbin (2016) reflects a growing consensus in economics, namely that non-cognitive skills are crucial determinants of meaningful life outcomes, including educational attainment, labour market success, health, and criminality – to name a few. The evidence from this body of work – termed the 'skill formation literature' or the 'economics of human development' – has strongly influenced the economics of education, where non-cognitive skills are increasingly regarded as both crucial inputs and outcomes of the schooling process. Consequently, non-cognitive skills have become central to education policy, with international development organisations and national education departments formally placing non-cognitive skills on their agendas (see for example Garcia (2014); OECD (2015); and Puerta and Valerio, (2016)).

Despite this surge of interest in non-cognitive skills among education researchers and policymakers alike, there is a dearth of evidence of the association between non-cognitive skills and learning outcomes from LMICs. An important unanswered question in the international literature, therefore, is whether the strong association between these skills and educational outcomes observed in HICs holds in contexts of severe resource deprivation that characterise much of the developing world. Evidence from the United States (US) suggests that non-cognitive skills, such as perseverance, positive attitudes toward school, and student engagement are strongly associated with academic success. But does it help to be perseverant in a school that does not have basic instructional materials? Is there any benefit to having a positive attitude toward mathematics when one's teacher lacks the content knowledge and pedagogical skill to effectively teach the curriculum? How is student engagement related to learning outcomes when there is limited instructional time and opportunity to learn in the school day?

These questions speak to Pritchett and Sandefur's (2013: 33) argument – which reflects an emerging perspective among other prominent development economists (see for example Ravallion (2020) and Hanushek (2021)) – for the importance of evaluating associations that are well-established in HICs in a broader range of contexts, particularly LMICs:

> "Inasmuch as development economics is a worthwhile, independent field of study – rather than purely a parasitic form of regional studies, applying the lessons of rich-country economies to poorer settings – its central conceit is that development is different. The economic, social, and institutional systems of poor countries operate differently in rich countries in ways that are sufficiently fundamental to require different models and different data."

As such, the next three chapters in this thesis aim to add evidence from South Africa to the international evidence base of the association between non-cognitive skills and student achievement. This is done with a view to improving our understanding of the role of non-cognitive skills in predicting learning outcomes not just in South Africa, but in other LMICs more broadly. It is hoped that this evidence will constitute a first step in a much larger research project that aims to add more evidence from LMICs to existing theories of the role of non-cognitive skills in education so that we may ultimately generate a superior understanding of how non-cognitive skills can support learning.

1.2. DEFINITIONAL ISSUES

In the economics literature, 'non-cognitive skills' is a rather opaque term used interchangeably with socio-emotional skills (Kraft, 2019) to refer to all the "personality traits, goals, motivations, and preferences that are valued in the labour market, in school, and in many other domains" (Heckman and Kautz, 2012: 2). Unsurprisingly, the use of the term 'non-cognitive skills' is the subject of ongoing debate in the literature. Two major points of contention are: (i) which term should be used to refer to these skills, with the current list including "behavioural skills, soft skills, personality traits, non-cognitive abilities, character, socio-emotional skills, and non-cognitive skills" (Garcia, 2014: 6), and (ii) the divide between cognitive and non-cognitive skills. Regarding the latter, a number of authors have argued that this is an artificial divide, since "every psychological process is cognitive in the sense of relying on processing information of some kind" (West *et al.*, 2016: 149). Researchers have also highlighted the lack of theoretical or empirical justifications for the different traits that are lumped under the umbrella of cognitive skills (for example, measured IQ and performance on specific tasks, such as naming letter sounds) on the one hand, and non-cognitive skills (such as personality traits and learnable skills, like time management) on the other (Duncan and Magnuson, 2011). These are major criticisms, and some authors have even advocated for the abolishment of the terms ‘cognitive’ and ‘non-cognitive’ skills from our vocabulary (for example Duncan and Magnuson (2011)).

I maintain, however, that these debates are largely due to the relative newness of the field, and while these issues around definitional clarity will undoubtedly have to be addressed in order to move forward in this field, doing so is beyond the scope of this thesis. I therefore echo Garcia’s (2014: 7) argument for the use of the term “non-cognitive skills”, namely that “it is still common practice to refer to the broad type or category of skill, and many key contributions in this area… use the term ‘non-cognitive skills’ rather than anything more specific.”

As such, the analyses in this thesis consider two quite distinct types of non-cognitive skills as inputs into the educational production process, namely student attitudes toward learning and school (Chapters 2 and 4) and ‘grit’ (Chapter 3). While these constructs are not usually studied together in the literature, they fall under the same broad label of ‘non-cognitive skills’ insofar as they meet the criteria for non-cognitive skills offered by Garcia (2014: 6), namely:

> “traits that are not directly represented by cognitive skills or by formal conceptual understanding, but instead by socio-emotional or behavioural characteristics that are not fixed traits of the personality, and that are linked to the educational process, either by being nurtured in the school years or by contributing to the development of cognitive skills (or both).”

It should be noted that the choice of non-cognitive skills studied in this thesis was informed largely by data availability. Unfortunately, a major challenge in examining the association between non-cognitive skills in learning outcomes in LMICs – and possibly part of the reason why there is a dearth of this evidence in the international literature – is that measures of these skills are rarely available in LMICs. While efforts are beginning to be made to measure these skills in LMICs (see for example Tovar García, (2017); Jukes *et al.* (2018); and He *et al.* (2021)), the availability of measures of non-cognitive skills is a major limitation in advancing the study of non-cognitive skills in LMIC contexts. The analyses in this thesis therefore makes use of measures of non-cognitive skills that are available for South Africa, in an attempt to show what can be learnt about the association between these skills and student achievement even with very limited measures of non-cognitive skills.

1.3. A BRIEF HISTORY OF NON-COGNITIVE SKILLS IN THE ECONOMICS OF EDUCATION

The idea that non-cognitive skills are important determinants of educational outcomes is, of course, not new. More than a century ago, psychologists Binet and Simon (1916: 254) noted that performance in school “admits other things than intelligence; to succeed in his studies, one must have qualities which depend on attention, will, and character." Despite this important caveat to the work of the architects of the first modern IQ test (Almlund *et al.*, 2011), education research would go on to focus on the development of cognitive skills as both a determinant and objective of formal schooling. A telling example of this is how Coleman's (1966) models of achievement in his *Equality of Educational Opportunity* report – which is widely considered to mark the inception of the economics of education as a discipline (Hanushek, 1979) – included non-cognitive skills as covariates. Moreover, Coleman found that internal locus of control had a stronger relationship with achievement than did all the school factors included in his model together. That this result was deliberately de-emphasised, both by Coleman and those who engaged with his work, illustrates just how little interest there was in non-cognitive skills within education research and policy at the time[1].

A decade later, economists Bowles and Gintis (1976) emphasised the role of non-cognitive skills in education, stressing non-cognitive traits over IQ in the inheritance of social class (Sampson, 2016). Bowles and Gintis were also the first to explicitly contrast non-cognitive skills with cognitive skills in determining educational outcomes (Farkas, 2011). Interestingly, the work of Bowles and Gintis did not immediately spur further interest among economists in the relationship between non-cognitive skills and educational outcomes. Instead, this research agenda was taken up by sociologists of education (beginning with the work of Jencks, Crouse and Mueser (1979)), who were especially concerned with expanding Bowles and Gintis' work on the links between non-cognitive skills, education, and social mobility. Education economists chose to focus on the relationship between school characteristics and achievement, maintaining that these factors are the most amenable to policy intervention (Hanushek, 1979). This perspective would persist in the economics of education for the next thirty years.

As the findings from labour economics (pioneered by Heckman and colleagues (2000, 2006) in the early 2000's) increasingly pointed to the importance of non-cognitive skills in determining meaningful life outcomes, however, education economists have re-visited these skills as potentially important inputs into the production of education. This shift is clearly illustrated in the fact that the first chapter of the fourth volume of the Handbook of the Economics of Education (Hanushek, Machin and Woessmann (eds), 2011) is titled *Personality Psychology and Economics* (Almlund *et al.*, 2011) and chronicles the history of the study of non-cognitive traits in economics research. These authors also document the substantial body of evidence of the strong link between non-cognitive skills and educational outcomes. Contributions to the literature on the role of non-cognitive skills in producing learning outcomes during the last decade have mainly comprised methodological contributions in the form of proposed statistical techniques for addressing problems of measurement error and reverse causality that plague earlier studies (see for example, Heckman and Corbin (2016); Johnes, Johnes and

[1] While the authors of the first studies in the economics of education acknowledged the importance of non-cognitive skills in determining learning outcomes, very little was known about the formation of non-cognitive skills at the time (including whether they could be fostered through intervention), and therefore these results were not considered relevant for policy purposes (see Hanushek (1968)).

López-Torres (2017); and Osher *et al.* (2018)). Attempts have also been made at evaluating the association between non-cognitive skills and achievement in experimental settings (Beattie, Laliberté and Oreopoulos, 2016; West *et al.*, 2016; Attanasio *et al.*, 2020). The most recent contributions to the literature on non-cognitive skills in the production of education have included attempts to formally unify the skill formation and educational production function research paradigms by developing complex econometric techniques for simultaneously estimating the effects of home and school environments on the production of cognitive and non-cognitive skills (Agostinelli, Saharkhiz and Wiswall, 2019). The results from these studies constitute strong empirical evidence of a causal link between non-cognitive skills and learning outcomes.

1.4. NON-COGNITIVE SKILLS AND EDUCATION IN LMICS

The evidence of the strong association between non-cognitive skills and meaningful life outcomes from HICs has sparked interest in these skills among the international development community. As Scorza *et al.* (2017: 1) argue:

> "Human development and economic development are intrinsically linked. Guided by human capital theory, economists are increasingly recognising the importance of a range of other skills - in addition to intelligence and technical skills – for economic success. Until fairly recently, years of education completed, literacy, numeracy, and IQ – often used as proxies for cognitive ability – were the main measures to assess the relationship between human capital and economic development. More recently, researchers and practitioners have acknowledged that skills such as the ability to work in groups, maintain good interpersonal relations and a positive attitude, control impulses and demonstrate goal-oriented behaviour are all critical to economic productivity and individual success."

Education researchers, in particular, are increasingly looking to interventions that target non-cognitive skills as having the potential to unlock learning in resource-limited countries. The evidence base for the role of non-cognitive skills in education in LMICs is extremely thin, however, with researchers only very recently beginning to address this gap in the literature (see for example Delprato, Akyeampong and Dunne (2017); Espinosa (2017); Aberra (2018); Kim, Brown and Weiss-Yagoda (2018); Miranda and Domingues (2018); Attanasio *et al.* (2020); and He *et al.* (2021)). The evidence presented in this thesis is intended to add to this evidence base by investigating the relationship between non-cognitive skills and learning outcomes in South Africa.

1.5. THE SOUTH AFRICAN CONTEXT

Even before the COVID-19 pandemic, learning was in crisis in South Africa. Standardised achievement tests consistently showed that the majority of South African children were failing to reach even basic proficiency in early grade literacy and numeracy, which form the foundation upon which all other academic skills are built (Howie *et al.*, 2017; Human Sciences Research Council, 2017; Zuze *et al.*, 2017). In addition to exceptionally low levels of aggregate performance, the country's education system is one of the most unequal in the world (Spaull, 2019), with present-day patterns of educational inequality largely reflecting the patterns of segregation and deprivation established during the apartheid era (Van der Berg, 2007; Branson and Lam, 2010; Spaull, 2013). Specifically, existing evidence points to a dualistic education system, where the poorest 80% of schools – those that served Coloured and Black children under apartheid - have remained largely dysfunctional under the democratic government (Spaull, 2014; Van der Berg *et al.*, 2016), while the wealthiest 20% of schools – mainly those that served White and Indian students under apartheid - are not dissimilar to schools in developed countries.

The local economics of education literature has hitherto focused on school-level and broader institutional factors in explaining these outcomes. Based on the evidence from this body of work, Van der Berg *et al.* (2016: 5) argue that there are four 'binding constraints' to improving educational outcomes in the country, namely; "(1) Weak institutional functionality, (2) Undue union influence, (3) Weak teacher content knowledge and pedagogical skill, and (4) Wasted learning time and insufficient opportunity to learn." Tragically, just when the South African education system was starting to show signs of progress (Van der Berg and Gustafsson, 2019; Gustafsson, 2020), the COVID-19 pandemic hit, resulting in the twin shocks of school closures and education budget cuts (Gustafsson and Nuga, 2020). Past research indicates that these shocks are likely to disrupt human capital accumulation in the country for years to come (Andrabi, Daniels and Das, 2020; World Bank Group, 2020).

The extreme circumstances that currently characterise the education landscape in South Africa make it all the more urgent to improve our understanding of the educational production process in the country, both for research and policy purposes. While the existing economics of education literature in South Africa has immensely enriched our knowledge of the role of school-level factors in predicting educational outcomes, there is much about the educational production process in South Africa that remains unexplained. The evidence from the international economics of education literature outlined above suggests that the role of non-cognitive skills in the educational production process may be an important piece of the puzzle, in terms of explaining learning outcomes which have not yet been considered in local research. Moreover, the existing evidence of large improvements in learning outcomes through interventions that raise non-cognitive skills from high-income countries (see Durlak *et al.* (2011) for a meta-analysis of this evidence) suggests that targeting these skills may be a particularly useful policy lever for raising learning outcomes in South Africa.

1.6. RESEARCH QUESTIONS

This thesis is composed of three empirical questions about the potential role of non-cognitive skills in predicting learning outcomes in South Africa. Each of the chapters in this thesis considers this question through a different lens, which informs the econometric strategy employed. Four datasets were employed in order to model the production of student achievement in two subjects (reading and mathematics) at different grade levels.

In Chapter 2, South African data from the Progress in International Reading Literacy Study (PIRLS) and the Trends in Mathematics and Science Study (TIMSS) are analysed through the lens of "academic resilience" to explore what makes some students achieve Grade 4 reading and Grade 9 mathematics results beyond expectations. The first research objective in this chapter is to identify academically resilient students in the PIRLS 2016 and TIMSS 2015 datasets. I consider how these students are distributed across schools of differing quality, and how they perform relative to the median student in their school. My second research objective is to determine ways in which these students differ systematically from their lower-achieving peers. Logistic ordinary least squares (OLS) regression analysis is used to investigate factors at the individual, family, and school level that contribute to the probability that a student will achieve exceptional academic results. In accordance with findings from the international literature, I find a strong association between student attitudes toward learning and school and the probability of exceptional reading performance in Grade 4 and mathematics performance in Grade 9 in South Africa. Similar to a number of existing studies, I find that the constructs aimed at capturing self-confidence, in particular, are strongly associated with the probability of academic resilience in both PIRLS and TIMSS.

Chapter 3 makes use of Grade 6 reading achievement data from a local study titled Leadership for Literacy to explore potential interaction effects between school functionality and the non-cognitive skill of 'grit' by estimating OLS regressions of achievement separately by school functionality and testing for interaction effects between grit and school functionality. The econometric analysis points to potential moderating effects between the perseverance subscale of grit and school functionality. This result makes an important contribution to the international literature on non-cognitive skills in education, since it is one of only a handful of studies that investigates whether other educational inputs interact with non-cognitive skills to predict learning outcomes, and the first to investigate whether school quality, specifically, interacts with the non-cognitive skill of grit. The results from Chapter 3 further suggest that that the nature of the interaction between perseverance and school functionality is not uniform across the distributions of these variables. In terms of the broader literature on potential interaction effects between non-cognitive skills and other educational inputs, this is an important result, since it provides empirical evidence for the theoretical possibility that the nature of interaction effects between non-cognitive skills and other educational inputs may vary at different points of the distributions of both non-cognitive skills and the inputs they interact with.

A natural extension of the results from Chapters 2 and 3 is to explore whether South Africa's gendered educational outcomes can be linked to gender differences in non-cognitive skills. The links between non-cognitive skills and gendered educational outcomes have been the subject of a number of studies from HICs (see for example Jacob (2002) and Duckworth and Seligman (2006)). This link especially deserves to be investigated in South Africa, as the country exhibits one of the largest pro-girl gaps in education in the world. This analysis is undertaken in Chapter 4. Again, PIRLS and TIMSS data is utilised to model student achievement, this time using the Grade 5 TIMSS data to model mathematics achievement. Oaxaca-Blinder decomposition analysis is used to investigate how much of South Africa's pro-girl achievement gap in these datasets can be explained by gender differences in observable characteristics, with a particular focus on the contribution of gender differences in non-cognitive skills. The results of the decomposition analysis provide some evidence that gender differences in non-cognitive skills are part of the reason for South Africa's gendered educational outcomes. However, these skills seem to be more predictive of the gender gap in Grade 4 reading achievement than Grade 5 mathematics achievement. The contributions of this chapter are twofold: (i) the results add to the evidence presented in Chapters 2 and 3 of a strong association between non-cognitive skills and achievement, and (ii) the results add more specificity to this evidence, showing that the contribution of non-cognitive skills to the pro-girl achievement gap is not uniform across subjects and grades. Overall, the analysis in this chapter illustrates clearly how focusing on non-cognitive skills as predictors of learning outcomes can enhance our understanding of hitherto unexplained features of South Africa's educational performance, such as the country's large and persistent pro-girl achievement gap.

1.7. CONTRIBUTIONS

This thesis makes an important contribution to the international evidence base of the association between non-cognitive skills and learning outcomes. This evidence base is currently dominated by evidence from HICs, which are not representative of the general human population. Given that the international community of economists of education are increasingly recognising that contextual factors may interact with educational inputs in the production of learning outcomes (Ravallion, 2020; Hanushek, 2021a), the dearth of evidence of the association between non-cognitive skills and learning outcomes from LMICs constitutes a major gap in the literature on non-cognitive skills in education. Given the attention currently afforded to non-cognitive skills in education research and practice internationally, it is imperative that more evidence from LMICs is introduced to the existing evidence base of the role of non-cognitive skills in education. This thesis aims to do so by presenting evidence of this association from South Africa. Although classified as an upper-middle income country, the extreme levels of inequality that characterise social contexts in the country mean many students in South Africa face schooling contexts that are not dissimilar to those in some of the poorest countries in the world. As such, the evidence provided in this thesis is relevant not only to the local economics of education literature, but also to the broader international literature on the role of non-cognitive skills in education in LMICs. Given the promising evidence from HICs of the impact of fostering non-cognitive skills on a broad range of meaningful life outcomes (Heckman, 2000), it is crucial from a development perspective that these skills are studied in LMICs. Although the evidence presented in this thesis is subject to a number of important limitations, the work contained in this thesis should be seen as a first step in advancing our understanding of the role of non-cognitive skills in producing learning outcomes in LMIC contexts.

CHAPTER 2: PERFORMANCE BEYOND EXPECTATIONS: EXAMINING CORRELATES OF EXCEPTIONAL ACADEMIC ACHIEVEMENT AMONG POOR STUDENTS IN SOUTH AFRICA

2.1. INTRODUCTION

Socio-economic status (SES) and educational outcomes are strongly linked across countries and education systems. However, a growing body of research documents the existence of students from disadvantaged socio-economic backgrounds who manage to achieve exceptional academic results. The literature on exceptional academic performance in high-poverty contexts is primarily concerned with identifying the individual and institutional factors that underpin this ability to achieve good academic results despite adverse socio-economic conditions. Given that an estimated 390 million poor children globally are at risk of achieving minimum proficiencies in learning (UNICEF, 2018) it is important to explore what enables a small minority of poor students to overcome the risk factors associated with poverty and achieve good academic results.

The present study, located in the South African context, uses data from PIRLS and TIMSS to explore the factors at the individual and institutional level that are associated with exceptional academic performance in the face of socio-economic disadvantage. The South African context is particularly suited to studying exceptional performance among socio-economically disadvantaged students for three reasons. Firstly, the country's average performance was among the worst in PIRLS 2016 (Foy *et al.*, 2016) and TIMSS 2015 (Mullis *et al.*, 2015). Secondly, South Africa's PIRLS and TIMSS scores are among the most unequal in the world, with SES being a major determinant of achievement. Lastly, even though South Africa is classified as an upper-middle-income country, income inequality is so pronounced that children on the lower end of the income distribution face severely under-resourced home and school environments that are akin to those faced by children in some of the poorest countries in the world (Case and Deaton, 2005). For these reasons, studying academic resilience in South Africa may provide crucial insights into the factors that are associated with academic success in contexts of poor average performance and severe resource constraints both at the home and school level.

The first research objective is to identify exceptional performers in the PIRLS 2016 and TIMSS 2015 datasets. I consider how these students are distributed across schools of differing quality, and how they perform relative to the median student in their school. This is done in order to determine whether the phenomenon whereby some students from poor backgrounds manage to achieve good results in PIRLS and TIMSS occurs at the level of the school (an 'outlier school' hypothesis), or at the level of the individual student (an 'outlier child' hypothesis). The second research objective explores the ways in which these students differ systematically from their lower-achieving peers. The analytical strategy aims to identify factors at the level of the individual and the school that are associated with unusually high results in the absence of crucial inputs, such as an affluent home background.

Contributing to a growing body of literature internationally which documents the existence of exceptional performers in large-scale educational assessments (Sandoval-Hernandez and Cortes, 2012; Erberber *et al.*, 2015; Sandoval-Hernandez and Bialowoski, 2016; Agasisti and Longobardi, 2017; Agasisti *et al.*, 2018), I find that exceptional performers in South Africa are scattered across schools of varying quality, a result which supports the 'outlier child' hypothesis. This result echoes findings from Wills (2017), Spaull and Pretorius (2019), and Wills and Hofmeyr (2019), and strongly suggests that there are important individual and home background characteristics that enable some students to achieve good PIRLS reading and TIMSS mathematics results, despite attending schools with very low average levels of performance. In particular, it is found that the constructs in the PIRLS and TIMSS data aimed at capturing student attitudes toward reading and mathematics are strong predictors of exceptional performance in these assessments, respectively. This finding supports evidence from Wills and Hofmeyr (2019) that non-cognitive skills are highly predictive of exceptional academic performance in South Africa, and adds to the growing international evidence base of a strong association between non-cognitive skills and exceptional academic achievement among students from disadvantaged socio-economic backgrounds.

2.2. LITERATURE REVIEW

2.2.1. THE STUDY OF ACADEMIC RESILIENCE

> "Resilient students may be a small minority, but they may also be crucial to our understanding of the characteristics and contexts that make a positive difference in the lives of vulnerable populations" (OECD, 2011: 14).

The study of resilience finds its roots in developmental psychology, where the term 'resilience' is used to refer to "the processes of, capacity for, or patterns of positive adaptation during or following exposure to adverse circumstances that have the potential to disrupt or destroy the successful functioning or development of the person" (Masten and Obradovic, 2008: 2). Economists of education have subsequently borrowed the term to describe socio-economically disadvantaged students who manage to overcome the odds and achieve good academic results (Sandoval-Hernandez and Bialowoski, 2016). The above quotation from the OECD's (2011) study of student resilience, entitled "Against the Odds: Disadvantaged Students Who Succeed in School" highlights the main motivation behind this line of enquiry, namely the idea that understanding what enables some students to achieve good results despite socio-economic disadvantage may help us inform policy that enables more students to do the same.

A finding that emerges from virtually all these studies is that non-cognitive skills may be an important part of the explanation for why some students manage to achieve good results despite challenging home and school contexts and low levels of aggregate performance (Borman and Rachuba, 2001; Cappella and Weinstein, 2001; Borman and Overman, 2004; OECD, 2011; Erberber *et al.*, 2015; Vera, Valenzuela and Sotomayor, 2015; Huang and Zhu, 2017; Agasisti *et al.*, 2018; Das, 2018). Student attitudes comprise a subcategory of non-cognitive skills that are often studied in this literature, since measures of student attitudes are widely available in the large educational assessment datasets used by economists of education. Student attitudes refer to the feelings and beliefs students have about learning and school, and are thought to play an important role in moderating student effort and motivation (Petscher, 2010).

2.2.2. LEARNER ENGAGEMENT

Engagement in reading and mathematics lessons is one dimension of student attitudes toward learning that has received increasing attention as a determinant of achievement in these subjects in recent years. Engagement is thought to influence achievement as a mediator of students' motivation and attention in these subjects (Tse and Xiao, 2014). A number of studies find evidence of a positive association between self-reported engagement in reading lessons and reading achievement (Connell, Spencer and Aber, 1994; Twist, Schagen and Hodgson, 2007; Tse and Xiao, 2014; Vera, Valenzuela and Sotomayor, 2015), as well as engagement in mathematics lessons and mathematics achievement (Connell, Spencer and Aber, 1994; Borman and Rachuba, 2001; Ma and Xu, 2004; Ma and Kishor, 2006; House and Telese, 2016). The issue of learner engagement in South African schools – particularly those serving low-income communities – has been studied most notably by Hoadley (2018: 4), who summarises twenty years of research on the topic as follows:

> "[Classroom exchanges between teachers and learners are characterised by] a communalizing approach to instruction, where… the teacher works with the whole class as a homogenous group, with little or no differentiation of tasks or attention to individual performances. This is contrasted with an individualizing pedagogy that emphasises more personalized relationships between teachers and learners."

This quotation suggests that South African teachers working in high-poverty schools do not emphasise individual performance or personalised relationships with individual learners. Given this feature of pedagogy in high-poverty schools in South Africa, it is important that we investigate whether there are students who achieve results that exceed the average levels of performance of their class, and, if such students can be identified, which factors are associated with individual exceptional performance.

2.2.3. SUBJECT-SPECIFIC ENJOYMENT

Enjoyment of reading and mathematics is another aspect of attitudes toward learning that is considered important for achievement in these subjects. Students who enjoy reading spend more time reading and have better concentration when reading (Malanchini *et al.*, 2017), which may account for the positive association between reading enjoyment and achievement found by a number of authors (see for example Wang and Guthrie, 2004; Twist, Schagen and Hodgson, 2007; Mol and Jolles, 2014; Tse and Xiao, 2014; McGeown *et al.*, 2015; and Malanchini *et al.*, 2017). Similarly, the positive association between enjoyment of mathematics and achievement (Reynolds and Walberg, 1992; Ma, 1997; Singh, Granville and Dika, 2002; Ma and Xu, 2004; Ma and Kishor, 2006; Zuze *et al.*, 2017) is often ascribed to increased involvement and attention during mathematics lessons, as well as more time spent on mathematics (Singh, Granville and Dika, 2002).

2.2.4. CONFIDENCE

Self-reported confidence in a subject is also thought to impact positively on achievement through its relationship with effort, where confidence is associated with a higher perceived payoff of exerting effort in a particular subject (Petscher, 2010). It is well-established in the economics of education literature that self-reported confidence in reading is positively associated with reading performance (House, 2003; Twist, Schagen and Hodgson, 2007; Park, 2011; Van de Gaer *et al.*, 2012; Ibourk, 2013; Retelsdorf, Köller and Möller, 2014; Tse and Xiao, 2014; McGeown *et al.*, 2015; Francis *et al.*, 2017), and that the same relationship exists between confidence in mathematics and mathematics achievement (Borman and Overman, 2004; Azina and Halimah, 2012; Abu-Hilal *et al.*, 2013; Miscevic-Kadijevic, 2015; House and Telese, 2016; West *et al.*, 2016). Notably, the aforementioned OECD (2011) study showed that even when augmenting academic achievement models to include individual student characteristics *and* school-level factors, self-confidence in science remains the most consistent predictor of exceptional science achievement among disadvantaged students in almost all countries participating in the Programme for International Student Assessment (PISA).

2.2.5. SCHOOL SAFETY

In addition to student attitudes toward specific subjects, student-perceived school safety has also received attention as an important determinant of achievement, although less so than subject-specific attitudes toward school. Two constructs measured in the PIRLS and TIMSS student background questionnaires are of relevance here, namely sense of school belonging (measured with questions such as "I feel safe at this school") and measures of the frequency of student bullying. There is a dearth of empirical evidence for the relationship between self-reported sense of belonging in school and exceptional achievement in reading and mathematics. The small number of existing studies on self-reported school belonging find a positive association between students' sense of belonging at school and achievement (Sari, 2012; Topçu, Erbilgin and Arikan, 2016; Thomson *et al.*, 2017). Bullying is more widely studied as a determinant of achievement, with existing studies finding evidence of a negative association between the frequency of student bullying and academic achievement (Ponzo, 2012; Cosgrove and Creaven, 2013; Tse and Xiao, 2014; Sandoval-Hernandez and Bialowoski, 2016; Topçu, Erbilgin and Arikan, 2016; Thomson *et al.*, 2017; Zuze *et al.*, 2017).

Given these findings of the importance of student attitudes for academic achievement, the analysis in this chapter seeks to determine which of these attitudes are predictive of exceptional academic performance among students from low socio-economic backgrounds in South Africa. Specifically, the analytical strategy aims to identify whether the student attitudes measured in the PIRLS and TIMSS student background questionnaires predict performance that exceeds expectations among socio-economically disadvantaged students.

2.3. DATA

Both PIRLS and TIMSS datasets are included in the analysis. This allows me to utilise information on academic performance in different subjects from students in different grades, with a view to investigating whether the same factors are associated with exceptional performance across literacy and mathematics, as well as for students in different grades.

2.3.1. PIRLS LITERACY 2016

PIRLS is an international large-scale literacy assessment conducted by the International Association of the Evaluation of Educational Achievement (IEA). PIRLS Literacy 2016 was administered by the Centre for Evaluation and Assessment (CEA) at the University of Pretoria. In addition to student assessment data, PIRLS collected contextual information from students, teachers, and school principals (Foy *et al.*, 2016). The school language policy of South Africa is currently implemented in such a way that the language of learning and teaching (LOLT) for the vast majority of students is their home language in Grades 1-3, and from Grade 4 there is a LOLT switch to English for the remaining school years (Spaull and Kotze, 2015). This means by Grade 4 the majority of South African students would have had limited exposure to English, and consequently PIRLS was administered in all of South Africa's 11 official languages. PIRLS 2016 employed a two-stage stratified cluster sampling design so that a nationally representative sample of schools was chosen according to province and the school's language of instruction in the foundation phase (Howie *et al.*, 2017). Within the sampled schools, classes were randomly selected for participation. Sampled classes thus make up the second-stage sampling units. All students in sampled classes present on the day of the assessment participated in the assessments. In 2016 the realised PIRLS sample consisted of 12,810 grade 4 students from 293 schools across South Africa. Unfortunately, there are significant proportions of missing data in PIRLS, especially in the home background questionnaire items. Since home background information is paramount in operationalising exceptional performance, missing data could not simply be dealt with using listwise deletion. Instead, a combination of different imputation methods and listwise deletion was used[2]. After imputation, the PIRLS sample consisted of 12,762 students. When identifying exceptional performers and exploring the factors associated with exceptional performance, the analysis was further limited to the poorest 75% of students in the PIRLS sample[3]. This resulted in a PIRLS sub-sample of 9,572 students.

---

[2] See Appendix A for a more detailed description of the imputation methods employed.

[3] See Section 2.5 for a description of the operationalization of SES in this chapter.

2.3.2. TIMSS 2015

TIMSS 2015[4] is also conducted by the IEA and was administered in South Africa by the Human Sciences Research Council (HSRC). TIMSS is only administered in English and Afrikaans, as students are expected to have fully made the switch to one of these languages by Grade 9. Although TIMSS consists of both mathematics and science assessments, only mathematics TIMSS scores were considered in my analysis of exceptional performance. TIMSS collected the same contextual information from students, parents, teachers and principal as PIRLS (Mullis *et al.*, 2015). Students were sampled using the same two-stage stratified cluster sampling design as employed in PIRLS. The TIMSS 2015 realised sample consisted of 12,514 Grade 9 students[5] from 292 schools. The same missing data concerns present in PIRLS plague the TIMSS data, thus the same combination of listwise deletion and imputation methods was used to deal with missing information[6]. After imputation, the TIMSS 2015 sample consisted of 12,419 students. When limiting the TIMSS sample to the poorest 75% of students for the analysis of exceptional performers, the resultant sub-sample consisted of 9,316 students.

2.4. SOUTH AFRICA'S OVERALL PERFORMANCE IN PIRLS AND TIMSS

South Africa's poor performance in PIRLS and TIMSS is well-documented (see for example Mullis *et al.*, 2015; Howie *et al.*, 2016). Figure 1 shows the distribution of reading scores for the full PIRLS sample of 12,762 students, and the distribution of mathematics scores for the full TIMSS sample of 12,419. The dotted lines represent PIRLS and TIMSS international benchmarks, at 400, 475, 550 and 625 representing the low, intermediate, high and advanced benchmarks, respectively.

The figure shows that the majority of participating students in PIRLS (79.8%) did not reach the low international benchmark (10,183 students). Only 774 students (6.1%) reached the intermediate international benchmark, 164 students (1.3%) reached the high international benchmark, and only 18 students (0.05%) reached the advanced international benchmark. The mean reading score for the country is 318, making South Africa the worst performer out of the 50 countries participating in PIRLS (Foy et al., 2016), with an overall PIRLS score lower than that of other upper-middle-income countries that participated in the assessment (Azerbaijan, Iran, and Kazakhstan), as well as lower-middle-income participating countries (Egypt, Georgia, and Morocco).

The picture is similar for South Africa's performance in the TIMSS mathematics assessment: 8,334 participating students (67.1%) did not reach the low international benchmark, and only 1,328 (10.7%)

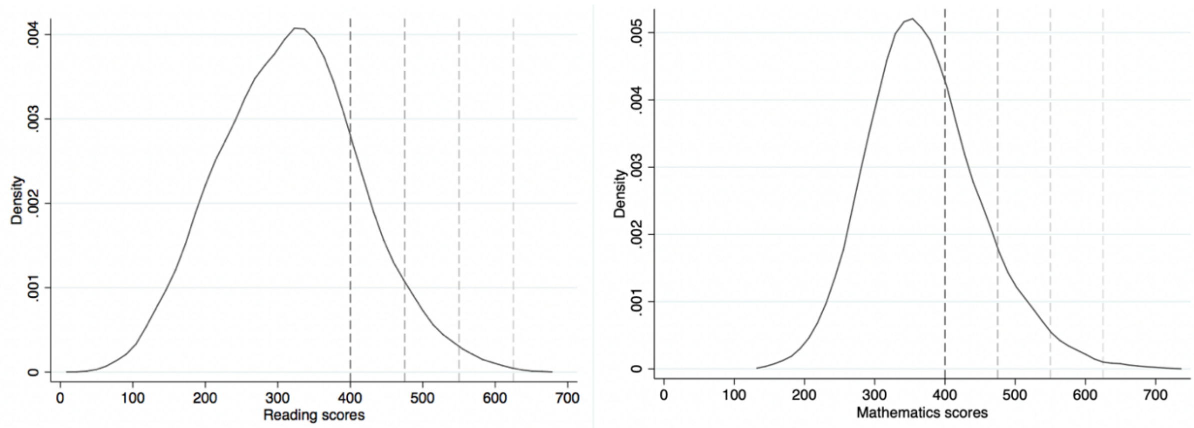
[4] Grade 9 TIMSS results are analysed in this chapter to investigate whether the same factors are associated with academic resilience in different subjects and at different grade levels (Grade 4 and Grade 9). Chapter 4 makes use of Grade 5 TIMSS results since the aim of that chapter is to compare gendered learning outcomes in reading and mathematics, respectively, for students who are similar in age (Grade 4 and Grade 5).

[5] Even though the TIMSS assessment is administered in Grade 8 in most countries, participating countries can choose to administer the assessment in Grade 9 if they suspect the assessment will be too difficult for Grade 8 students (Mullis *et al.*, 2015).

[6] See Appendix A for a more detailed description of the imputation methods employed.

reached the intermediate international benchmark. Only 324 students (2.6%) reached the high international benchmark, and 61 students (0.5%) reached the advanced international benchmark of 625 points. South Africa achieved the second-lowest mathematics score in TIMSS 2015, at 372 points (Mullis *et al.*, 2015). As is the case for South Africa's performance in PIRLS, this overall performance is lower than that of other participating upper-middle-income countries (Botswana, Iran, Jordan, Kazakhstan, Lebanon, and Thailand) and lower-middle-countries (Egypt, Georgia, and Morocco) that participated in the TIMSS assessment.

Figure 1: Distribution of South Africa's reading (Grade 4) and mathematics (Grade 9) scores

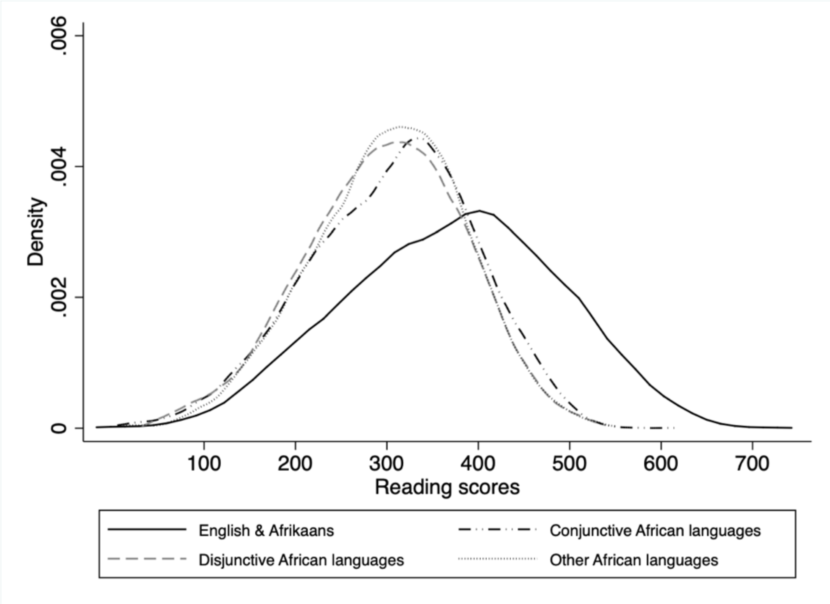


Notes: Kernel density distribution using epanechnikov, bandwidth = 13.4619; Sources: PIRLS 2016 and TIMSS 2015

It is a remnant of the apartheid schooling system that South Africa's inequality in reading scores at the primary school level is strongly related to language of learning and teaching (LOLT). The apartheid education system consisted of multiple racially defined departments of education, and education policies deliberately aimed to deliver inferior quality education to Black students compared to White students (Spaull, 2013; Van der Berg, 2008). Most schools with African languages as the LOLT would have been part of the Bantu education department during apartheid, and many of these schools still suffer the inertia of decades of limited resources and poor management (Van der Berg et al, 2011; Spaull, 2015). In this sense, LOLT can be considered a proxy for school disadvantage. The impact of this on school quality can be seen in Figure 2, which shows distributions of PIRLS reading scores by the language the test was written in. African languages were grouped according to their orthography, namely conjunctive (isiXhosa, Siswati, isiNdebele, isiZulu) and disjunctive (Sepedi, Setswana, Sesotho) orthographies (Spaull, Pretorius and Mohohlwane, 2018). Tshivenda and Xitsonga are classified as "other" African languages since they are not part of either of these language families.

It is clear from the figure that students writing the test in an African language performed much worse in PIRLS than students writing in English or Afrikaans. Given these different performance distributions by LOLT, and the fact that a school's LOLT overlaps with contextual factors, such as neighbourhood poverty and school resources, we might expect differences in the factors associated with exceptional performance based on the LOLT of the school. This question is explored in Section 2.6.

Figure 2: Distribution of PIRLS reading scores by test language (Grade 4)

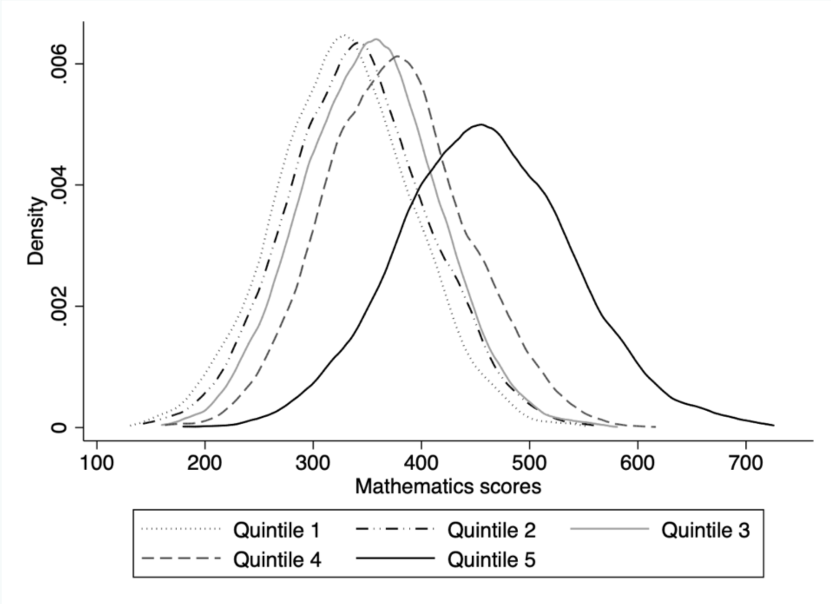


Notes: Kernel density distribution using epanechnikov, bandwidth = 13.4619; Source: PIRLS 2016

The same inequality in schooling outcomes exists at the high school level in South Africa, however in TIMSS, language of instruction cannot be used as a proxy for apartheid era education departments. School disadvantage in TIMSS is therefore measured as the SES of the school, categorised as five school quintiles[7], with the poorest 80% of students in schools categorised as Quintile 1-4, and the wealthiest 20% of students attending Quintile 5 schools. Figure 3 shows the distribution of TIMSS mathematics scores by school quintile and illustrates that a similar bimodal distribution exists, whereby the distribution of mathematics scores in Quintile 5 schools lies clearly to the right of mathematics scores in the bottom four school quintiles. Once again, given these differences in overall mathematics scores by school quintiles, we might expect differences in the factors that are associated with exceptional performance in Quintile 1-4 schools compared to Quintile 5 schools. This is explored in Section 2.7.

[7] School quintiles were constructed using TIMSS data on student SES. School SES is measured as the mean student SES at the school level.

Figure 3: Distribution of TIMSS mathematics scores by school quintile (Grade 9)



Notes: Kernel density distribution using epanechnikov, bandwidth = 11.8405; Source: TIMSS 2015

2.5. IDENTIFYING EXCEPTIONAL PERFORMERS

Exceptional performance is defined in this thesis as academic achievement that exceeds socio-demographic expectations. Following Borman and Overman (2004), characteristics of students' home environments are used to define the "expectations" relative to students' performance, in order to identify exceptional performers. This decision was informed by two considerations. First, given the strong association between student SES and academic achievement (internationally and especially in South Africa), SES and other home-background factors were used which may proxy for wealth to set expectations for academic achievement. Secondly, by using home-level variables to define exceptional performance (the first stage of the analytical strategy), modelling the probability of being an exceptional performer on individual- and school-level factors could be undertaken in the second stage of the analytical strategy. This is advantageous since student attitudes and school characteristics are more amenable to changes in policy and practice than home background factors. This allows me to identify factors that are associated with better academic achievement in low SES contexts which are more amenable to change through policy or practice.

To operationalise this definition, it is necessary to first obtain a predicted test score based on socio-economic background factors. This is done by regressing reading scores on student SES for the PIRLS sub-sample and mathematics scores on SES for the TIMSS sub-sample, as per the following equation:

$$Y_{is} = \beta_0 + \beta_1 SES_{is} + \varepsilon_{is} \qquad (1)$$

Here Y_{is} represents either the reading or mathematics score of the ith student in schools. SES_{is} is a vector of the students' home background characteristics and its square, which is an index derived from both parents' education, an index of home possessions, and a binary variable indicating whether the student attends a school in a township or remote rural area or not[8], using principal components analysis (PCA). The residuals of the equation, ε_{is} (the difference between the actual value of Y and the predicted value of Y), represent the part of an individual student's reading or mathematics score that cannot be explained by these socio-economic factors.

To identify students whose test scores exceed expectations, it is necessary to specify some level of residual test performance above the predicted score as per equation (1). This level is set at 1.5 standard deviations above the mean residual reading score for the full PIRLS and TIMSS samples[9]. To limit the sample of exceptional performers to students from poor socio-economic backgrounds, a further constraint is added, such that exceptional performers must be drawn from the bottom 75% of the asset index distributions in the PIRLS and TIMSS samples, respectively. Exceptional performers in the succeeding analysis are therefore students with asset index scores at or below the 75th percentile whose test scores lie 1.5 standard deviations above the mean, after accounting for the socio-economic background factors listed above.

Using this definition, I identify 553 exceptional performers in the PIRLS sample of 12,762, that is, 4.33% of the total sample of students. These students are distributed across 191 schools, that is, 71.27% of schools in the sample. Similarly, this operationalization of exceptional performance yields 555 exceptional performers in the TIMSS sample of 12,419 (4.47%), also distributed across 191 schools (69.0% of schools in the sample). It must be noted that, as is the case with any self-reported measures of SES, the SES is measured with error. The measurement error may be particularly large in the South African PIRLS and TIMSS data given that many missing values in the student background questionnaires had to be imputed. A limitation of this study is, therefore, that some exceptional performers may be misidentified.

Figure 4 and Figure 5 provide graphical representations of the reading and mathematics scores, plotted against asset index scores, for exceptional performers in PIRLS and TIMSS, respectively. Exceptional performers in African language schools and Quintile 1-4 schools in PIRLS and TIMSS, respectively, are represented by the crosses. Just over half (53.16%) of the exceptional performers in PIRLS attend African language schools (294 students), and 61.96% of exceptional performers in TIMSS (347 students) are drawn from Quintile 1-4 schools. The dotted lines represent the 75th percentile on the asset

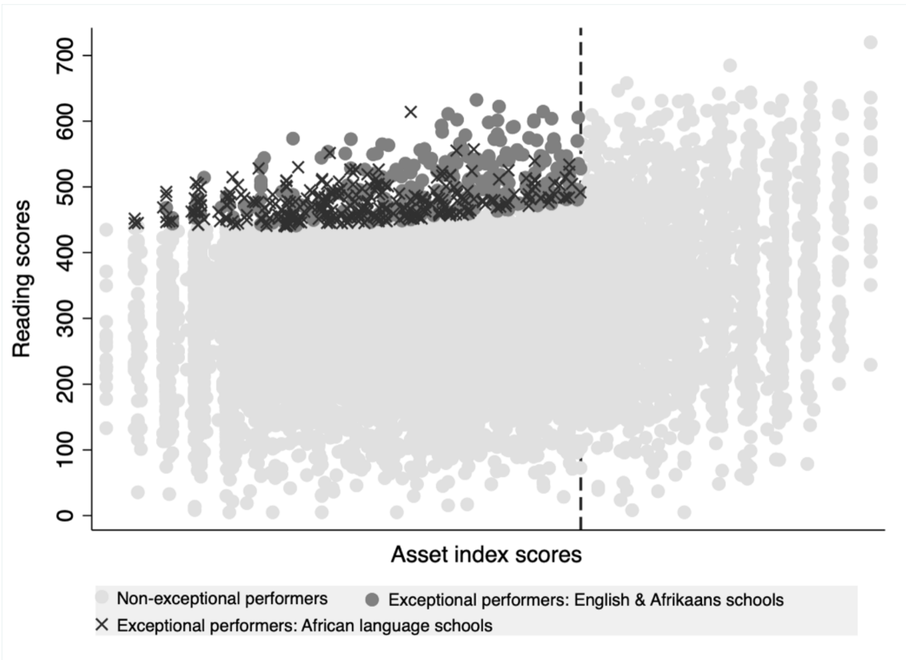
[8] PIRLS does not collect direct information about whether students live in an urban or rural area. Given that urban/rural is an important sociodemographic divide in South Africa, it was decided to include information about the location of the school in the SES measure. Even though the school location variable captures information at the school level, and not the home, 70-80% of South African students in rural and township schools walk to school (Grant, 2014), indicating that they live in the same geographic area as their schools. In this sense, school location can be considered a crude proxy for the type of geographic area of students' homes.

[9] That is, the samples that result after imputing missing data on key variables.

index in each sample, that is, the cut-off point that is used to limit both samples in the multivariate analysis.

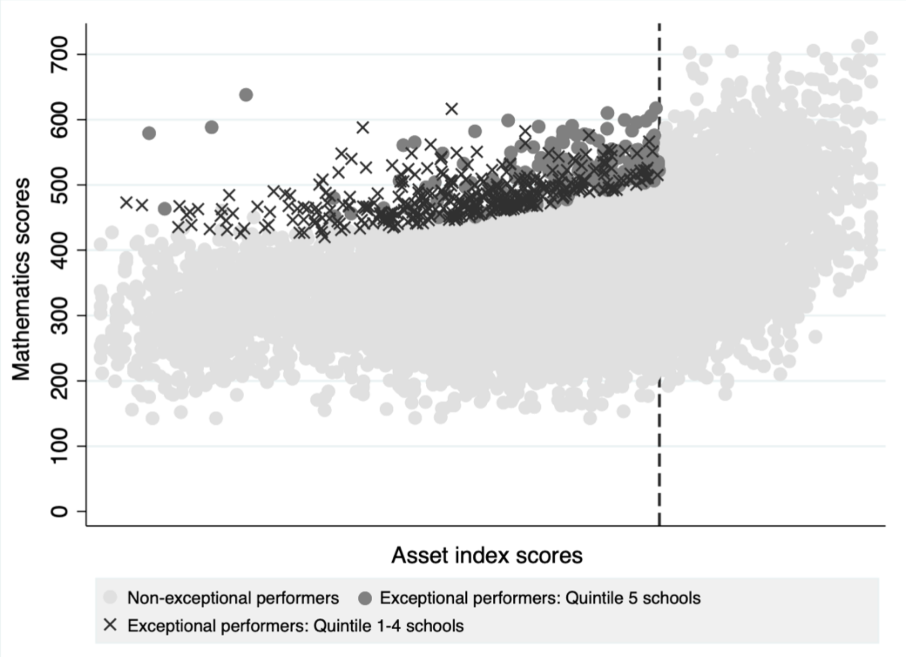
The figures show that exceptional performers in both PIRLS and TIMSS achieved test scores above the low international benchmark (400 points). In PIRLS, reaching the low international benchmark amounts to having basic reading skills, in other words students could retrieve explicitly stated information and make straightforward inferences when reading the less difficult texts (Mullis *et al.*, 2017). In TIMSS, reaching the low international benchmark means students have an elementary knowledge of whole numbers and basic graphs (Mullis *et al.*, 2015). The median reading score among exceptional performers in PIRLS is 482 points, that is, above the intermediate international benchmark of 475 points. At this level of reading proficiency, students are able to integrate and interpret story events and information, in addition to making basic inferences (Mullis *et al.*, 2017). At 484 points, the median mathematics score among exceptional performers in TIMSS is also above the intermediate international benchmark. At this level of achievement, students "can apply basic mathematical knowledge in a variety of situations" (Mullis *et al.*, 2015).

Figure 4: Identifying exceptional performers in PIRLS 2016 (Grade 4)



Notes: The dotted vertical line represents the 75th percentile on the asset index. Source: own estimations from PIRLS 2016.

Figure 5: Identifying exceptional performers in TIMSS 2015 (Grade 9)



Notes: The dotted vertical line represents the 75th percentile on the asset index. Source: own estimations from TIMSS 2015.

Figure 6 provides a graphical representation of the reading scores of exceptional performers in PIRLS, relative to the median reading score in their school. The crosses represent the reading scores of exceptional performers in African language schools, while the dark grey dots represent the reading scores of exceptional performers in English and Afrikaans schools. It is clear from the figures that there are exceptional performers even in schools with very poor average performance. In addition, exceptional performers in worse performing schools far outperformed the median student in their school, while the gap between exceptional performers' test scores and their school's median test score is much smaller for exceptional performers drawn from better performing schools. Indeed, Figure 6 shows that there are a number of exceptional performers in better performing schools whose reading scores were below that of the median student in their school[10]. This points to the important result that exceptional performers, as operationalized here, are not simply the top achievers in their schools. Rather, exceptional performers are those students whose reading scores exceed expectations, given their SES. Figure 7 shows similar patterns in the mathematics achievement of exceptional performers in TIMSS.

[10] There are 17 exceptional performers in PIRLS (3.07%) and 36 exceptional performers in TIMSS (6.49%) whose assessment scores are equal to or below the median score of their school.

Figure 6: Distribution of exceptional performers by school (PIRLS 2016)

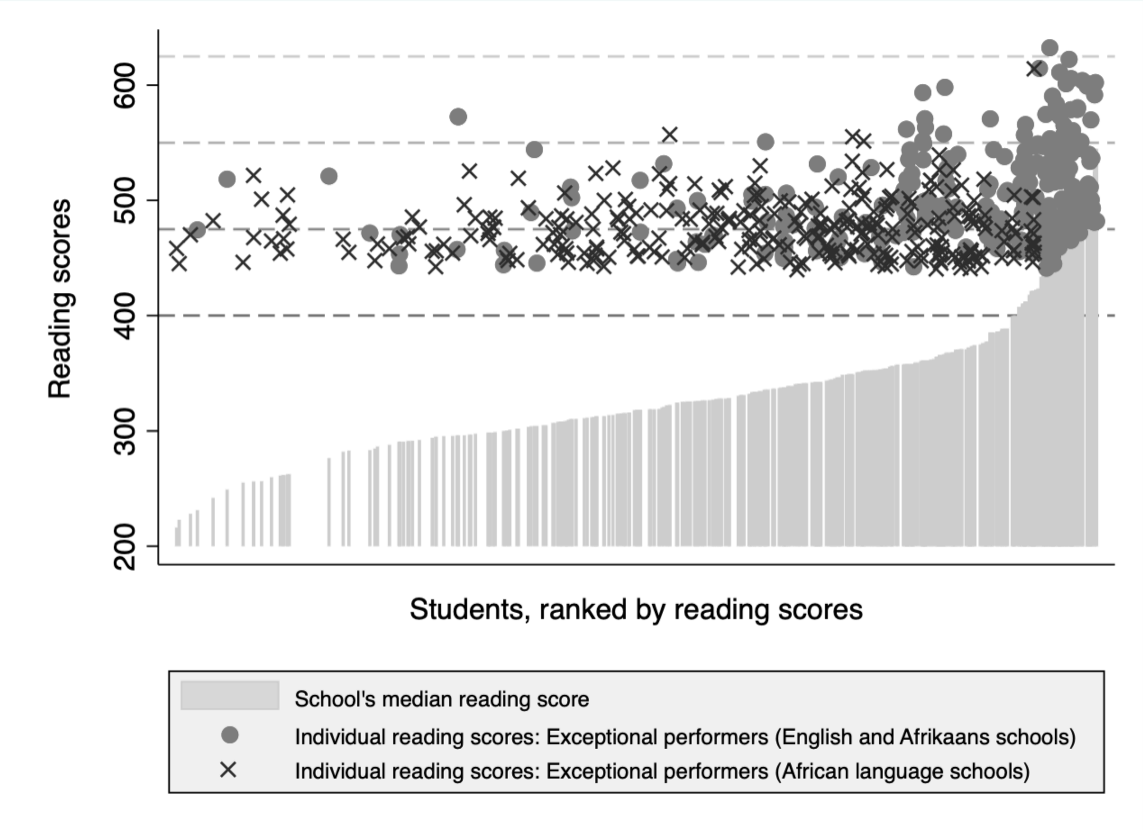


Figure 7: Distribution of exceptional performers by school (TIMSS 2015)

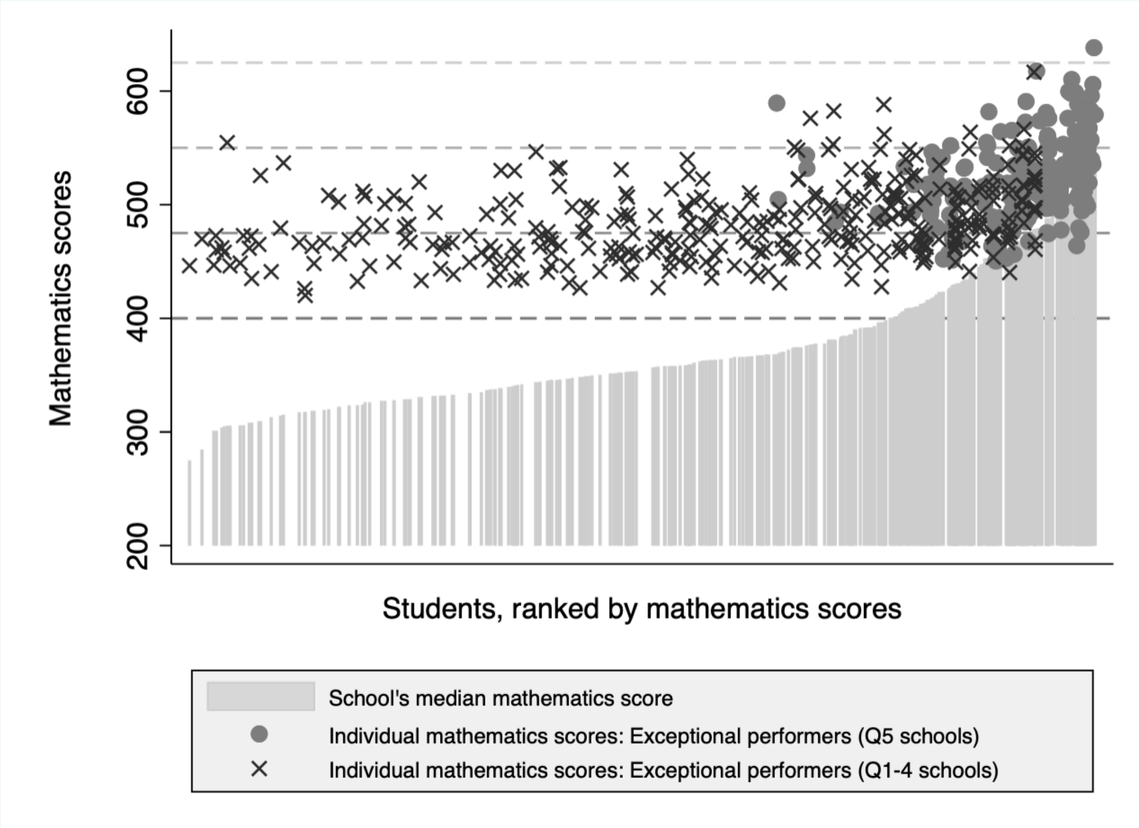


Table 1 shows descriptive differences in individual and school characteristics between exceptional and non-exceptional performers in the PIRLS and TIMSS data, respectively. The results in the table show that exceptional performers exhibit very different individual characteristics compared to their peers, in both the PIRLS and TIMSS sub-samples. Exceptional performers in PIRLS are younger, more likely to be girls, absent less often, experience bullying less often, have a higher sense of school belonging, report being more engaged in their reading lessons and have more confidence in reading than their peers. The only counterintuitive result in Table 1 is the lower score on the reading for enjoyment index for exceptional performers, relative to their peers. Possible reasons for this result are explored in Section 2.7.

The schools attended by exceptional versus non-exceptional performers in PIRLS also differ significantly along a number of dimensions. In terms of LOLT, African language exceptional performers are underrepresented relative to the proportion of students in African language schools in the PIRLS sub-sample: only 53% of exceptional performers are drawn from African language schools, while this proportion is 81% for the wider PIRLS sample. Schools producing exceptional performers appear to have more physical resources than schools without exceptional performers: these schools are more likely to have a library and, at least, one computer. Interestingly, exceptional performers are not significantly more likely to be taught by teachers who have at least a Bachelor's degree[11]. In general, the proportions of Grade 4 language teachers who have at least a Bachelor's degree in this sub-sample of PIRLS is very low, at 35% for teachers of students in the comparison group, and an even lower 33% for teachers of exceptional performers. School composition indicates the average student SES, which is the mean asset index[12] score at the school level. According to this measure, exceptional performers attend schools whose school bodies are wealthier, on average, than schools who did not produce any exceptional performers.

As is the case for PIRLS, exceptional performers in TIMSS are also younger than the comparison group of students. However, the difference in mean age between exceptional performers and the comparison group is much larger than in PIRLS. Exceptional performers in TIMSS are around 9 months younger than the comparison group, on average, while exceptional performers in PIRLS are only about 3 months younger than their peers in the comparison group. The larger age gap observed for the Grade 9 TIMSS sample may be attributed to the cumulative effects of repetition over time, where overage learners are more prevalent in later grades (Branson, Hofmeyr and Lam, 2014; Van Wyk, 2015).

Exceptional performers in TIMSS are also absent less often, experience bullying less often, have a higher sense of school belonging, report being more engaged in their mathematics lessons, and have more confidence in mathematics, compared to their peers. Unlike exceptional performers in PIRLS who do not report enjoying reading more than their counterparts, those in TIMSS report enjoying mathematics more than their peers, a more intuitive result. The results in Table 1 further indicate that girls are no more likely to be exceptional performers than boys in TIMSS. This is not altogether surprising since there were no significant gender differences in mathematics achievement for age appropriate students in the full South African TIMSS sample in 2015 (Zuze *et al.*, 2017).

The schools attended by exceptional performers also differ from the rest of the schools in the TIMSS sub-sample. These schools are less likely to be Quintile 1-4 schools, and more likely to have at least one computer. Interestingly, schools producing exceptional performers in TIMSS are not significantly more likely to have a library, and, as is the case in PIRLS, exceptional performers in TIMSS are not more likely to be taught by a teacher with at least a Bachelor's degree than non-exceptional performers.

[11] It must be noted that having a Bachelor's degree is age-related among South African teachers, since the qualifications framework for the teaching profession changed in 1996 so that a Bachelor's degree became a requirement for teachers entering the profession (Sayed, 2002).

[12] See Section 2.5 for a description of the construction of this index.

Table 1: Descriptive statistics: Exceptional versus non-exceptional performers

	PIRLS 2016 (Grade 4)		TIMSS 2015 (Grade 9)	
	Exceptional performers	Non-exceptional performers	Exceptional performers	Non-exceptional performers
Individual characteristics				
Female	0.71*** (0.02)	0.49 (0.01)	0.54 (0.02)	0.53 (0.01)
Age	10.00*** (0.03)	10.23 (0.01)	14.68*** (0.04)	15.45 (0.01)
Often absent	1.75*** (0.05)	2.26 (0.01)	1.28*** (0.03)	1.72 (0.01)
Frequency with which test language is spoken at home	0.87 (0.01)	0.90 (0.00)	0.43*** (0.02)	0.28 (0.00)
School safety & attitude indices				
Bullying index	2.02*** (0.03)	2.30 (0.01)	1.68*** (0.03)	1.85 (0.01)
School belonging index	2.60*** (0.02)	2.47 (0.01)	2.62*** (0.02)	2.55 (0.01)
Enjoyment index	1.31*** (0.02)	1.83 (0.02)	3.56*** (0.03)	3.29 (0.01)
Engagement index	2.76*** (0.02)	2.51 (0.01)	2.57*** (0.03)	2.53 (0.01)
Confidence index	2.25*** (0.03)	1.64 (0.01)	1.99*** (0.03)	1.54 (0.01)
School characteristics				
African language school (PIRLS) / Quintile 1-4 school (TIMSS)	0.53*** (0.02)	0.81 (0.01)	0.64*** (0.02)	0.92 0.00
School has a library	0.40*** (0.02)	0.29 (0.00)	0.43 (0.02)	0.40 (0.01)
School has at least one computer	0.37*** (0.02)	0.24 (0.00)	0.57*** (0.02)	0.41 (0.01)
Teacher has at least a Bachelor’s degree	0.33 (0.02)	0.35 (0.01)	0.96 (0.01)	0.96 (0.00)
SES of the school body	0.63*** (0.05)	-0.04 (0.01)	-0.04*** (0.04)	0.00 (0.01)
Mean test scores	490*** (1.58)	297 (0.90)	498*** (2.44)	346 (3.05)
School median	369*** (2.62)	308 (0.52)	417*** (2.43)	351 (3.22)
Observations	553	9019	555	8761

Notes: Standard errors are in parentheses. Statistically significant at the following levels: *p<0.1, **p<0.05, ***p<0.01.

2.6. DIFFERENCES BETWEEN EXCEPTIONAL PERFORMERS BY SCHOOL TYPE

Given the achievement differences in PIRLS between English and Afrikaans schools compared to African language schools highlighted in Section 2.4, we might expect exceptional performers to look very different, depending on whether they attend African language schools or English and Afrikaans schools. Table 2 shows differences in mean scores on a number of individual characteristics between exceptional performers who wrote the test in an African language and exceptional performers who wrote the test in English or Afrikaans. The results in the table suggest that there are indeed important differences between these two groups of exceptional performers. Firstly, the mean reading score of exceptional performers in English and Afrikaans schools is higher than that of exceptional performers in African language schools (546 compared with 498). However, it should be kept in mind that the median reading performance of African language schools producing exceptional performers is almost a whole standard deviation (100 points) lower than the median reading scores of English and Afrikaans schools producing outliers, at 342 and 434, respectively.

This points to the finding that differences in exceptional performers in African language schools' reading scores and the median reading score of their class are much larger than differences in reading scores between exceptional performers in English and Afrikaans schools and their school's median reading score. On average, exceptional performers in African language schools achieved reading scores 138 points higher than the median performer in their school, whereas exceptional performers in English and Afrikaans schools achieved scores 101 points higher than the median performer in their school.

Table 2 also shows that exceptional performers in African language schools are slightly younger, have a higher sense of school belonging[13], and are more engaged in reading than exceptional performers in English and Afrikaans schools. Interestingly, exceptional performers in African language schools scored lower on the index measuring reading enjoyment than those in English and Afrikaans schools. There are no statistically significant differences in gender, student absenteeism, mean scores on the bullying index, and mean levels of confidence in reading between exceptional performers in African language schools and their counterparts in English and Afrikaans schools.

As expected, African language schools producing exceptional performers differ significantly from those whose LOLT is English or Afrikaans. These African language schools are less likely to have a library or at least one computer, and exceptional performers in these schools are less likely to have a language teacher who has at least a Bachelor's degree. The school bodies of African language schools are also much poorer than those of English or Afrikaans schools producing exceptional performers.

Similar differences are observed between exceptional performers in TIMSS who attend Quintile 1-4 schools and their counterparts in Quintile 5 schools (Table 2). At 384 points, the median mathematics score of Quintile 1-4 schools producing exceptional performers is almost 100 points (one standard deviation) lower than that of Quintile 5 schools producing exceptional performers (479 points). The mean difference between the mathematics scores of exceptional performers and the median student in their school is much larger in Quintile 1-4 schools than in Quintile 5 schools, with exceptional performers in Quintile 1-4 schools scoring almost an entire standard deviation (95 points) higher than

[13] See Appendix A for a description of this and other student-level variables included in the multivariate analysis.

the median mathematics scores in their school, while Quintile 5 exceptional performers only scored 37 points higher than the median student in their school, on average.

There are also interesting differences in individual-level student characteristics between exceptional performers in Quintile 1-4 schools and those in Quintile 5 schools in TIMSS. As is the case for exceptional performers in African language schools in PIRLS, exceptional performers in Quintile 1-4 schools in TIMSS also have a higher sense of school belonging and report being more engaged in their mathematics lessons. Unlike exceptional performers in African language schools in PIRLS, however, those in Quintile 1-4 schools in TIMSS report enjoying mathematics more than their counterparts in Quintile 5 schools and scored higher on the index measuring confidence in mathematics. Interestingly, Quintile 1-4 exceptional performers reported being bullied more often than those in Quintile 5 schools.

Table 2 also shows that Quintile 1-4 schools producing exceptional performers differ from Quintile 5 schools producing exceptional performers, in terms of physical resources, where the former are less likely to have a school library and much less likely to have computers. There is no statistically significant difference in the proportion of exceptional performers in Quintile 1-4 schools whose mathematics teachers have at least a Bachelor's degree compared to exceptional performers in Quintile 5 schools.

Table 2: Differences between exceptional performers drawn from different types of schools in PIRLS and TIMSS

	PIRLS 2016 (Grade 4)		TIMSS 2015 (Grade 9)	
	African language	English & Afrikaans	Quintile 1-4	Quintile 5
Individual characteristics				
Female	0.74 (0.03)	0.68 (0.03)	0.53 (0.03)	0.58 (0.03)
Age	9.94** (0.05)	10.07 (0.05)	15.06 (0.05)	14.98 (0.05)
Often absent	1.76 (0.07)	1.75 (0.07)	2.28 (0.04)	2.26 (0.05)
Frequency with which test language is spoken at home	0.92*** (0.02)	0.81 (0.02)	0.36*** (0.03)	0.57 (0.03)
School safety and attitude indices				
Bullying index	2.05 (0.05)	1.99 (0.05)	1.71*** (0.03)	1.54 (0.04)
School belonging index	3.69*** (0.03)	2.51 (0.04)	3.70*** (0.03)	3.46 (0.04)
Enjoyment index	1.20*** (0.02)	1.42 (0.04)	3.66*** (0.03)	3.22 (0.05)
Engagement index	2.84*** (0.02)	2.68 (0.03)	3.69*** (0.03)	3.43 (0.05)
Confidence index	2.27 (0.04)	2.24 (0.05)	3.07*** (0.04)	2.89 (0.05)
School characteristics				
School has a library	0.30*** (0.03)	0.51 (0.03)	0.40** (0.03)	0.49 (0.04)
School has at least one computer	0.31*** (0.03)	0.44 (0.03)	0.48*** (0.03)	0.72 (0.03)
Teacher has at least a Bachelor's degree	0.24*** (0.03)	0.44 (0.03)	0.97 (0.01)	0.94 (0.02)
SES of school body	-0.06*** (0.05)	1.41 (0.07)	-0.35*** (0.05)	0.51 (0.05)
Mean test scores	337*** (1.46)	404 (2.60)	468*** (2.44)	514 (3.05)
School median	342*** (2.13)	434 (4.03)	384*** (2.43)	479 (3.22)
Observations	294	259	347	213

Notes: Standard errors are reported in parentheses. Statistically significant at the following levels: *p<0.1, **p<0.05, ***p<0.01.

2.7. MULTIVARIATE ANALYSIS

2.7.1. ESTIMATING THE PROBABILITY OF EXCEPTIONAL PERFORMANCE

Given these descriptive differences between exceptional performers and their peers, I consider whether these associations remain significant in a multivariate context. To do this, I follow the well-established and generally accepted convention in the literature (Sandoval-Hernandez and Bialowoski, 2016), of using a logistic regression model to estimate the probability of being an exceptional performer as a function of a number of a defined set of predictors.

The results from the logistic regressions are presented in Table 3. Models 1, 2, and 3 estimate the probability of exceptional performance in PIRLS, while Models 4, 5, and 6 estimate the probability of exceptional performance in TIMSS. Coefficients are presented as odds ratios which reflect the odds of being an exceptional performer associated with a given score on the independent variable, relative to the reference category. For example, the coefficient of 1.975 on "Female" in the second column of Table 3 indicates that girls are 1.975 times as likely as boys to be exceptional performers. Coefficient values smaller than one indicate a negative association between a given covariate and the probability of being an exceptional performer. For example, the coefficient of 0.724 on the covariate measuring the frequency of student absenteeism in the second column of Table 3 indicates that being absent once every two weeks is associated with a 28% decrease in the probability of being an exceptional performer compared with the reference category, that is, being absent never or almost never.

(I) INDIVIDUAL CHARACTERISTICS

The logistic regression results presented in Table 3 echo the descriptive differences between exceptional performers and their peers described in Section 2.6. Girls are almost twice as likely as boys to be exceptional performers when controlling for only individual factors (Model 1). The coefficient on the female dummy remains significant in both Models 2 and 3, when student attitudes and school characteristics, respectively, are controlled for. This result is consistent with the fact that girls consistently outperform boys in reading in South Africa (Van Broekhuizen and Spaull, 2017), and with existing studies that find a gender gap in favour of girls when estimating the covariates of exceptional performance in reading (Finn and Rock, 1997; Cappella and Weinstein, 2001; Vera, Valenzuela and Sotomayor, 2015; Wills and Hofmeyr, 2019) among socio-economically disadvantaged students. This pro-girl gap in reading outcomes is investigated further in Chapter 4 of this thesis. Interestingly, the coefficient on the female dummy becomes smaller when school safety and student attitude measures are added to the logistic regression (Model 2), suggesting that these measures are correlated with gender in PIRLS. A correlation matrix of the explanatory variables can be found in Appendix A (Table A3), and shows that gender is indeed correlated with all of the student attitude variables included in Models 2 and 3. The potential role of gender differences in student attitudes in explaining gender differences in PIRLS reading scores is investigated explicitly in Chapter 4 of this thesis.

The analysis in Chapter 4 also points to a pro-girl gap in Grade 5 mathematics achievement, however the results presented in Table 3 suggest there is no clear pro-girl advantage in Grade 9 mathematics achievement, with the female dummy remaining a non-significant predictor of exceptional performance in the multivariate context. Interestingly, this is the case even when controlling only for individual factors in TIMSS (Model 4), suggesting that gender and student attitudes are not correlated in TIMSS. This can further be seen in the correlation matrix of the explanatory variables in TIMSS (Table A4 of Appendix A). As is the case with PIRLS, this result is investigated further in Chapter 4 of this thesis, however Chapter 4 makes use of Grade 5 TIMSS data (and not Grade 9). The non-significance of the coefficient on the female dummy in TIMSS adds to existing evidence of no gender differences in exceptional mathematics achievement among socio-economically disadvantaged students in the high school grades (Sandoval-Hernandez and Bialowoski, 2016; Agasisti et al., 2018).

Age remains significantly associated with exceptional performance in a multivariate context, with older students being less likely to be exceptional performers in both PIRLS and TIMSS. Interestingly, the odds ratio associated with age becomes slightly smaller when adding student attitude and school safety measures to the regression (Models 2 and 5), an effect which is more pronounced in PIRLS (Model 2). This suggests age is also correlated with student attitudes, which is again shown in the correlation matrices in Table A3 and Table A4 of Appendix A. The fact that age remains a significant predictor of exceptional performance even when adding student attitudes and school characteristics to the logistic regressions suggests an independent effect of student age on the probability of exceptional performance. This is especially the case in TIMSS, where even when controlling for covariates at both the individual and school level (Model 6), being one year older is associated with being 40% less likely (odds ratio of 0.596) to be an exceptional performer. The association between age and exceptional performance is less than half as big in PIRLS, with being one year older being associated with a 14% decrease in the probability of being an exceptional performer.

Similarly, self-reported student absenteeism maintains its significance as a predictor of exceptional performance when controlling for other covariates, with more frequent absenteeism being negatively associated with the probability of being an exceptional performer. This supports existing evidence of a negative association between student absenteeism and academic achievement in South Africa (Van der Berg and Louw, 2006; Shepherd, 2011).

The frequency with which the language of the test is spoken at home also emerges as a significant predictor of exceptional performance in five out of the six logistic regressions. Here, the response categories were collapsed into a binary variable, with students who reported speaking the language of the test at home "almost" or "almost always" scoring one, and those who reported speaking it "sometimes" or "never" scoring zero. Interestingly, this variable loses its significance when controlling for school characteristics in PIRLS (Model 3). This may be due to the fact that many students in this sub-sample of PIRLS (89%) report "always" or "almost always" speaking the language of the test at home, which reflects the fact that PIRLS 2016 was administered in the school's LOLT in the foundation phase. In TIMSS, the proportion of students who "always" or "almost always" speak the language of the test at home is only 29%, reflecting the fact that English or Afrikaans is not the home language of the vast majority of students in this sub-sample of TIMSS. The results in Table 3 suggest students who are taught in their home language have an advantage in terms of the probability of being exceptional performers, with these students being 1.93 times as likely to be identified as such, compared to their peers who reported speaking the test language at home “sometimes” or “never”.

(II) STUDENT ATTITUDES

Students’ confidence in reading in PIRLS and mathematics in TIMSS emerge as the student attitude indices with the strongest associations with exceptional performance. In PIRLS, being confident in reading is associated with being 2.35 times more likely to be an exceptional performer, relative to students who are “not confident” in reading. Similarly, students who are confident in mathematics in TIMSS are 2.27 times more likely to be identified as exceptional performers than students who are “not confident” in mathematics in the full model specification in TIMSS (Model 6). This result echoes Stankov and Lee's (2014) conclusion, based on a review of five studies, that measures of self-confidence have the highest predictive power in studies that use large-scale assessment data to examine the relationship between self-reported confidence in students’ academic abilities and academic achievement.

Subject-specific student engagement is only significantly associated with academic resilience in the full model specification in PIRLS, and not in TIMSS. This is an interesting result in light of existing findings of the importance of student engagement for mathematics achievement, and provides some support for the notion that while engagement is associated with achievement in general, it is not a significant predictor of exceptional mathematics achievement for socio-economically disadvantaged students.

Students’ self-reported sense of belonging at school is also only associated with exceptional performance in PIRLS. This student attitude index is not significantly associated with exceptional performance in TIMSS, which is a surprising result given existing studies that find a positive association between school belonging and achievement. Self-reported frequency of experiencing bullying is negatively associated with exceptional performance in three out of the four basic regressions. It is interesting that bullying is not significantly associated with the probability of being an exceptional performer in the full specification (Model 4) in TIMSS, given that other studies find a relationship with student bullying and achievement.

(III) SCHOOL CHARACTERISTICS

Of the two physical school resources controlled for in Models 3 and 6, only attending a school with at least one computer is significantly associated with the probability of exceptional performance in both PIRLS and TIMSS. If we interpret having a computer as a proxy for physical school resources, this result suggests students in better resourced schools are more likely to be identified as exceptional performers. Attending a school with a library is not significantly associated with the probability of exceptional performance in either PIRLS or TIMSS, after controlling for other factors. The coefficients on teacher qualifications reflect the descriptive statistics in Table 1, where there are no statistically significant differences between exceptional performers and their peers, in terms of the proportions of students whose teachers have obtained at least a Bachelor's degree.

Table 3: Logistic regressions of exceptional performance: Odds ratios

	PIRLS 2016 (Grade 4)			TIMSS 2015 (Grade 9)		
	(1) PIRLS	(2) PIRLS	(3) PIRLS	(4) TIMSS	(5) TIMSS	(6) TIMSS
Individual characteristics						
Female	1.975*** (4.99)	1.833*** (4.73)	1.729*** (4.30)	0.890 (0.90)	0.996 (0.03)	0.929 (0.58)
Age	0.773*** (4.00)	0.890*** (1.61)	0.859** (2.04)	0.546*** (8.04)	0.583*** (6.86)	0.596*** (6.65)
Often absent	0.724*** (5.75)	0.810*** (4.23)	0.831*** (3.69)	0.608*** (5.50)	0.643*** (5.04)	0.652*** (4.76)
Test language spoken at home: Always or almost always	0.680*** (2.64)	0.634** (2.27)	1.070 (0.34)	2.353*** (5.01)	2.273*** (5.09)	1.929*** (4.45)
School safety and attitude indices						
Confidence in reading (PIRLS) / mathematics (TIMSS)		2.348*** (10.07)	2.131*** (9.73)		2.266*** (6.21)	2.372*** (6.31)
Engagement in reading (PIRLS) / mathematics (TIMSS) lessons		1.347** (2.19)	1.336** (2.12)		0.804* (1.87)	0.858 (1.41)
Enjoyment of reading (PIRLS) / mathematics (TIMSS)		1.028 (0.18)	1.137 (0.92)		1.058 (0.54)	1.275** (2.43)
Sense of school belonging		0.977 (0.22)	1.190* (1.73)		0.926 (0.44)	1.113 (0.71)
Frequency of experiencing bullying		0.815*** (2.74)	0.775*** (3.53)		0.811** (2.25)	0.927 (0.81)
School characteristics						
School library			1.293 (1.39)			0.965 (0.16)
School has at least one computer			1.466** (2.11)			1.870*** (2.73)
Teacher has at least a Bachelor's degree			0.857 (0.89)			1.480 (1.15)
SES of school body			1.515*** (5.05)			1.558* (1.95)
Observations	9,572	9,572	9,572	9,316	9,316	9,316

Notes: Correlation matrices of the explanatory variables are reported in Tables A3 and A4 of Appendix A. Coefficients expressed in odds ratios. Standard errors are in parentheses and clustered at the school level. Asterisks indicate statistical significance levels at *p<0.1; **p<0.05; ***p<0.01.

2.7.2. DIFFERENCES IN PREDICTORS OF EXCEPTIONAL PERFORMANCE BY SCHOOL TYPE

Given the descriptive differences between exceptional performers in African language schools compared with exceptional performers in English and Afrikaans schools highlighted in Section 2.6, it is likely that different covariates are associated with the probability of exceptional performance for these two groups of students, respectively. To investigate this hypothesis, the PIRLS sub-sample was divided into students who wrote the test in an African language and those who wrote the test in English or Afrikaans and the probability of exceptional performance for each of these subpopulations was estimated separately. The results from these estimations are presented in Table 4.

Comparing the coefficients of Models 7 and 8 illuminates important differences between the predictors of exceptional performance in African language schools compared to English and Afrikaans schools. First, the coefficient on being female, although still positive, is slightly smaller for English and Afrikaans schools (1.609 compared with 1.865 for African language schools). This suggests that girls in African language schools have a more pronounced advantage in terms of the probability of being exceptional performers than girls in English and Afrikaans school, a result that points to a potential intersection between gender and SES in producing learning outcomes learning outcomes, something that is explicitly investigated in Chapter 4 of this thesis.

The association between the frequency with which the language of the test is spoken at home also differs by the test language, with this variable not being significantly associated with the probability of exceptional performance among students who wrote the test in an African language. By contrast, students in English or Afrikaans schools who indicated that they "always" or "almost always" spoke the language of the test at home are 1.590 times more likely to be identified as exceptional performers, relative to students who "sometimes" or "never" spoke the test language at home.

In terms of student attitudes, self-reported confidence in reading remains a large and significant predictor of exceptional performance in Models 7 and 8, however the size of the coefficients also differ by school type. Specifically, being confident in reading is associated with a higher probability of exceptional performance among students who wrote the test in English or Afrikaans, relative to students who wrote the test in an African language. Self-reported engagement in reading lessons remains an non-significant predictor of academic resilience when splitting the sample by LOLT of the school. By contrast, the coefficient on reading enjoyment gains significance when limiting the sample to students in African language schools. These results point to potential interaction effects between non-cognitive skills such as student attitudes and school characteristics in determining learning outcomes. This possibility is investigated in the next chapter of this thesis.

The association between students' self-reported sense of belonging at school also differs by the LOLT of the school. Interestingly, the frequency with which students experience bullying at school loses significance as a predictor of exceptional performance in Models 7 and 8.

As is the case with exceptional performers in schools with different languages of instruction in PIRLS, there are significant descriptive differences between exceptional performers in Quintile 1-4 schools and those in Quintile 5 schools in TIMSS (Table 2). Once again, one might therefore expect different factors to be associated with exceptional performance in these different types of schools. To investigate this hypothesis, a similar strategy was employed with the TIMSS data as with the PIRLS data in Models 7 and 8, where the sub-sample was divided into students in Quintile 1-4 schools and Quintile 5 schools, respectively, and the probability of exceptional performance for each of these subpopulations was estimated separately. The results of these estimations are also presented in Table 4.

The results in the table show that gender remains non-significant as a predictor of exceptional performance in both Quintile 1-4 and Quintile 5 schools in TIMSS. The coefficients on age remain very similar to those in Model 6, with older students still being less likely to be exceptional performers. The coefficients on the confidence index in TIMSS also mirror those in PIRLS: self-reported confidence in mathematics remains a significant predictor of exceptional performance when splitting the TIMSS sample into Quintile 1-4 and Quintile 5 schools, respectively. Although the coefficient on the variable measuring confidence in mathematics is slightly larger in Model 10 (that is, when the regression is limited to Quintile 5 schools), a Wald test of significance indicates that the coefficients on the confidence index are not statistically significantly different between Models 9 and 10. This suggests the strength of the association between confidence in mathematics and the probability of exceptional performance is similar for students in Quintile 1-4 schools compared to students in Quintile 5 schools.

Self-reported student engagement in mathematics lessons remains an insignificant predictor of academic resilience when splitting the TIMSS sample by school type, as is the case for PIRLS. Interestingly, mirroring the PIRLS results for African language schools, the enjoyment index gains significance when limiting the sample to Quintile 1-4 schools. Again, this points to variation in the association between student attitudes and achievement by school type, which is explored further in the next chapter.

It is interesting to note that the only school-level covariate which retains its significance when splitting the TIMSS sub-sample in this way is the dummy indicating whether a school has at least one computer. As is the case for the African language sub-sample in PIRLS, this variable is positively associated with exceptional performance in Quintile 1-4 schools. The coefficients on the dummy variables indicating whether schools have a library and whether teachers have obtained at least a Bachelor's degree remain not significant in Models 9 and 10.

Table 4: Logistic regressions of exceptional performance, separately by school type: Odds ratios

	PIRLS 2016 (Grade 4)		TIMSS 2015 (Grade 9)	
	(7) African language	(8) English & Afrikaans	(9) Quintile 1-4	(10) Quintile 5
Individual characteristics				
Female	1.865***	1.609*	0.813	1.134
	(4.18)	(1.93)	(1.33)	(0.63)
Age	0.777***	0.686**	0.570***	0.654***
	(3.22)	(2.62)	(6.07)	(2.68)
Often absent	0.854*	0.758***	0.667***	0.599***
	(1.91)	(3.00)	(3.72)	(3.74)
Test language spoken at home:	1.022	1.590***	2.016***	1.755***
Always or almost always	(0.10)	(3.60)	(4.25)	(1.78)
School safety and attitude indices				
Confidence in reading (PIRLS) /	1.974***	2.410***	2.318***	2.480***
mathematics (TIMSS)	(5.94)	(9.56)	(5.86)	(3.13)
Engagement in reading (PIRLS) /	1.026	1.098	0.806	0.882
mathematics (TIMSS) lessons	(0.09)	(0.52)	(1.51)	(0.77)
Enjoyment of reading (PIRLS) /	1.882***	1.004	1.719***	0.973
mathematics (TIMSS)	(2.96)	(0.02)	(4.39)	(0.14)
Sense of school belonging	1.322	1.188*	1.190	1.029
	(1.64)	(1.80)	(1.04)	(0.12)
Frequency of experiencing bullying	0.848	0.844	1.059	0.788
	(1.65)	(1.47)	(0.69)	(1.20)
School characteristics				
School library	1.102	1.123	0.831	1.182
	(0.35)	(0.40)	(0.85)	(0.38)
School has at least one computer	1.525*	1.547	1.711**	2.221
	(1.78)	(1.62)	(2.46)	(1.54)
Teacher has at least a	0.884	1.043	0.993	1.604
Bachelor's degree	(0.48)	(0.19)	(0.02)	(0.94)
SES of school body	0.874	2.432***	2.217***	0.903
	(1.25)	(7.86)	(3.07)	(0.26)
Observations	7,544	2,028	8,428	888

Coefficients expressed in odds ratios. Standard errors are in parentheses and clustered at the school level. Asterisks indicate statistical significance levels at *p<0.1; **p<0.05; ***p<0.01.

2.8. ROBUSTNESS CHECKS

Given that the choice of cut-off points for defining exceptional performance is somewhat arbitrary, it is necessary to test the robustness of the results of the main estimation against different cut-off points for defining exceptional performers. The results of using different cut-off points are presented in Table 5 below. The models in Table 5 include the full set of covariates included in Models 3 and 6 above, but only the coefficients on the main variables of interest (school safety and student attitude indices) are reported in the table. In models 11 and 12, exceptional performers are defined as students with asset index scores at or below the 70th percentile and residual test scores 1.5 standard deviations or more above the mean residual test score in the in the first-stage estimation. In models 13 and 14, exceptional performers are defined as students with asset index scores at or below the 70th percentile and residual test scores one standard deviation or more above the mean residual test score in the first-stage estimation. The results in the table show that the coefficients on the school safety and student attitude measures remain largely unchanged when using these different cut-off points, except for the student engagement index in PIRLS, which loses significance. Importantly, the main result from Section 2.7 – that confidence in reading and mathematics is the strongest predictor of exceptional performance in

PIRLS and TIMSS, respectively – is robust to the use of these different cut-off points to define exceptional performance.

Table 5: Coefficients on school safety and attitude indices using different cut-off points to define exceptional performance (Odds ratios)

	1.5 SD's above mean, 70th perc. on asset index		1 SD above mean, 70th perc. on asset index	
	(11) PIRLS	(12) TIMSS	(13) PIRLS	(14) TIMSS
School safety and attitude indices				
Confidence in reading (PIRLS) / mathematics (TIMSS)	2.175*** (9.49)	2.263*** (7.92)	1.958*** (10.87)	2.172*** (9.48)
Engagement in reading (PIRLS) / mathematics (TIMSS) lessons	1.058 (0.36)	0.887 (1.01)	1.130 (1.08)	0.924 (0.80)
Enjoyment of reading (PIRLS) / mathematics (TIMSS)	1.284* (1.78)	1.119 (1.18)	1.254** (2.22)	1.088 (0.93)
Sense of school belonging	1.255** (2.43)	0.862 (0.92)	1.256*** (3.00)	0.967 (0.025)
Frequency of experiencing bullying	0.793*** (2.88)	0.803*** (2.76)	0.762*** (4.88)	0.788*** (3.09)
Observations	8,967	9,316	8,967	9,316

Notes: Models include controls for individual and school characteristics, but are not reported here. Coefficients expressed in odds ratios. Standard errors are in parentheses and clustered at the school level. Asterisks indicate statistical significance levels at *p<0.1; **p<0.05; ***p<0.01.

Another way to test the validity of the constructs in the multivariate analysis is to show whether the main relationships of interest exist along a continuum – that is, to examine the relationships between the school safety and attitude indices and the probability of being an average performer or a below-average performer, as opposed to an exceptional performer. To do this, I create a categorical variable of resilience, where students who achieved residual test scores in the first estimation 1.5 standard deviations below the mean and lower are defined as "below average performers" and assigned a value of -1. Students who achieved residual test scores 1.5 standard deviations about the mean are defined as "average performers" and assigned a value of 0. The same operationalisation of exceptional performance as in the main estimation is used, whereby exceptional performers are defined as students who achieved residual test scores 1.5 or more standard deviations above the mean, and are assigned a value of 1. In the second stage estimations, I use multinomial logistic least-squares estimation, this time to estimate the probability of being a below-average performer or exceptional performer, relative to being an average performer. The results of this multinomial logistic regression are reported in Table 6 below. The coefficients on each of the covariates in models 1 and 2 represent the relative risk ratios of moving into the "below average performer" category from the "average performer" category, while the coefficients in models 3 and 4 represent the relative risk ratios of moving into "exceptional performer" category from the "average performer" category. Positive values indicate an increase in the multinomial log-odds of moving into a different performance category associated with a one-unit increase in the covariates, and negative values indicate that a one-unit increase in the explanatory variable is associated with a reduction in the multinomial log-odds of moving into a different performance category from the "average performer" category.

The results in Table 6 provide additional support for the construct validity of both academic resilience and the school safety and student attitude measures included in the estimation, particularly the variable measuring student confidence. In PIRLS a one-unit increase on the student confidence index is associated with a 0.428 unit decrease in the relative risk ratio of being a below-average performer, and a 0.747 unit increase in the relative risk ratio of being an exceptional performer, relative to being an average performer. In TIMSS, a one-unit increase on the student confidence index is associated with a 0.403 unit decrease in the probability of being a below-average performer, and a 0.806 unit increase in the probability of being an exceptional performer, relative to being an average performer. The fact that the relationships in the main estimation (Table 3) exist along a continuum lends further support to the validity of academic resilience as well as the covariates included in the main estimation.

Table 6: Multinomial logistic regressions of performance (Relative risk ratios)

	Below Average Performer		Exceptional Performer	
	(1) PIRLS	(2) TIMSS	(3) PIRLS	(4) TIMSS
Individual characteristics				
Female	-0.896***	0.167*	0.500***	-0.090
	(7.70)	(1.65)	(3.74)	(0.78)
Age	0.232***	0.431***	-0.260***	-0.565***
	(2.86)	(9.75)	(3.46)	(8.00)
Often absent	0.121***	0.216***	-0.219***	-0.406***
	(3.12)	(5.02)	(3.73)	(5.05)
Test language spoken at home: Always or almost always	-0.052 (0.43)	-0.440*** (2.98)	0.420*** (3.25)	0.781*** (5.52)
School safety and attitude indices				
Confidence in reading (PIRLS) / mathematics (TIMSS)	-0.428*** (4.52)	-0.403*** (3.81)	0.747*** (9.59)	0.806*** (7.87)
Engagement in reading (PIRLS) / mathematics (TIMSS) lessons	-0.202** (2.01)	-0.104 (1.10)	0.062 (0.42)	-0.124 (1.04)
Enjoyment of reading (PIRLS) / mathematics (TIMSS)	-0.388*** (4.47)	-0.331*** (3.97)	0.165 (1.29)	0.099 (1.03)
Sense of school belonging	-0.192**	-0.129	0.187**	-0.152
	(2.49)	(1.10)	(2.17)	(0.94)
Frequency of experiencing bullying	0.252***	0.444***	-0.208***	-0.204**
	(3.16)	(5.07)	(2.73)	(2.58)
School characteristics				
School library	0.043	-0.219	0.070	-0.004
	(0.20)	(1.23)	(0.38)	(0.01)
School has at least one computer	-0.158	-0.345*	0.431**	0.854***
	(0.73)	(1.90)	(2.41)	(3.46)
Teacher has at least a Bachelor’s degree	0.286 (1.51)	0.230 (0.40)	0.032 (0.18)	0.030 (0.10)
SES of school body	-0.097	0.538***	0.414***	-0.248***
	(0.69)	(7.14)	(4.40)	(3.36)
Constant	-3.122***	-8.776***	-2.894***	5.367***
	(3.33)	(10.28)	(3.03)	(4.31)
Observations	9,572	9,316	9,572	9,316

Notes: Coefficients expressed as relative risk ratios. The reference category for the dependent variable is “average performer”. Standard errors are in parentheses and clustered at the school level. Asterisks indicate statistical significance levels at *p<0.1; **p<0.05; ***p<0.01.

2.9. MEASUREMENT CHALLENGES

One major limitation of the analysis presented here is that the associations reported are purely correlational. Even if we assume an association is causal, it is not possible to determine the direction of causality with the type of analysis presented here. This is particularly problematic since the direction of causality may be reversed, where students with better academic skills have better attitudes towards reading and mathematics, as measured by the school belonging, enjoyment, engagement and confidence indices. Existing studies that explore the relationship between student attitudes and academic achievement have produced different theoretical models, such as the skill-development model and the self-enhancement model (Abu-Hilal *et al.*, 2013). According to the skill-development model, attitudes toward a domain (such as reading or mathematics) is the result of achievement in that domain. The self-enhancement model proposes the opposite direction of causality, whereby attitudes about a domain are the primary cause of achievement in that domain (Abu-Hilal *et al.*, 2013). Existing research has not been able to provide evidence as to which of these models better explains the observed relationship between student attitudes and achievement (Wang and Lin, 2008). While quasi-experimental techniques such as instrumental variable analysis could theoretically be used to identify whether the skill-development model holds over and above the self-enhancement model (Cordero, Cristóbal and Santín, 2018), the PIRLS and TIMSS data do not contain variables that could be used as valid instruments for estimating the association between student attitudes and achievement. Some authors have suggested using students' perceptions of the value of certain subjects such as mathematics as instruments (see for example Gamboa, Rodríguez Acosta & García-Suaza (2013)), however serious concerns remain about the extent to which data from student responses can be used as instruments (Cordero, Cristóbal & Santín, 2018). This limits the possibility of using instrumental variable analysis to deal with endogeneity in the association between student attitudes and achievement, a limitation that plagues all studies which explore this association using cross-sectional large-scale assessment data.

Another measurement concern relates to the validity of self-report questionnaires as measures of student attitudes. Self-report questionnaires constitute the most common approach to assessing student attitudes, such as sense of belonging, student engagement, and self-confidence. However, as Duckworth and Yeager (2015) argue, responding to such questionnaire items requires a complex process of reflection on the part of students. Of particular concern for interpreting the results presented in Table 3 is that the very first stage of this process – reading and understanding questionnaire items – requires literacy. Given South Africa's poor overall results in PIRLS especially, it is worth trying to disentangle the effect of literacy from responses to student attitude items.

Of particular concern among respondents with low literacy is what Marsh (1984) calls negative items bias, which he defines as occurring "when a child responds inappropriately by saying 'true' to a negative statement when his or her responses to positive items have consistently indicated that the opposite response would be more appropriate, or vice versa" (Marsh, 1984: 37). Weems, Onwuegbuzie and Collins (2009) argue that negatively worded items are particularly difficult for poor readers to answer, since negative ideas occupy twice as much space in working memory as positive ideas. This presents a bigger problem for poor readers, since individuals with low reading ability experience semantic processing problems that limit their ability to generate inferences while engaged in the reading process (Weems, Onwuegbuzie and Collins, 2009). In light of these findings, a number of authors concur that negatively worded statements add confusion that results in such items "measuring the students' ability to read carefully rather than their objective-based skills" (Carey, 2001: 126).

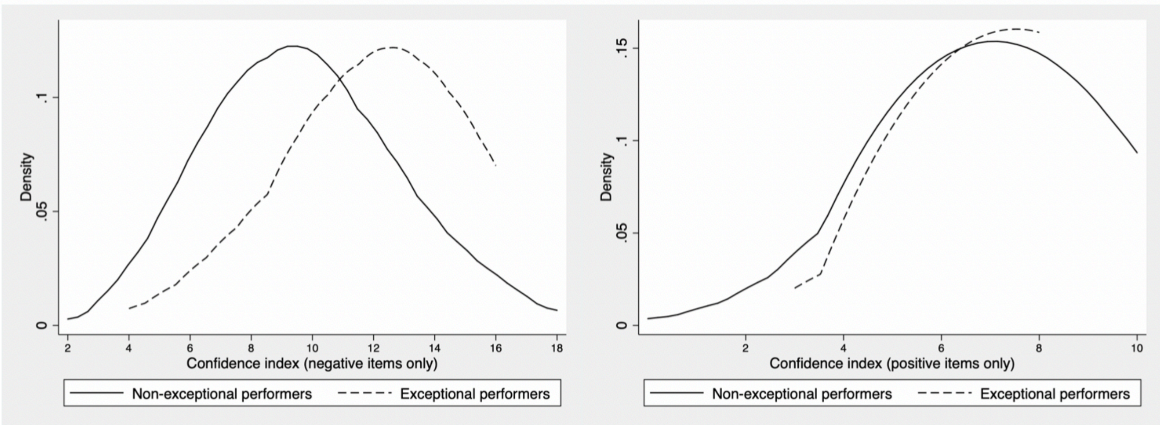
The index measuring confidence in reading is one of the only student attitude indices derived from the PIRLS and TIMSS student background questionnaires that includes negatively worded items, and the only attitude index that consists predominantly of negatively worded items. Given widespread concern about the validity of responses to negatively worded items in surveys, the fact that the only scale in PIRLS and TIMSS comprised predominantly of negatively worded items emerges as one of the strongest predictors of exceptional performance in reading and mathematics requires further investigation.

This situation is exacerbated by the fact that the potential effect of low literacy and that of low real levels of confidence in reading are indistinguishable. Students with low levels of literacy could respond "Agree a lot" to the statement "I am bad at reading" either because they understand the content of the question and know they struggle with reading, or because they do not understand the content of the question and answered "Agree a lot" to all the questions. Thus, it is difficult to establish whether answers on items related to confidence in reading, especially negatively-worded items, are a true reflection of students' perceptions of their reading ability (i.e. their confidence in reading), or simply due to their low levels of literacy. The same potential problem plagues the confidence in mathematics index in TIMSS 2015. Although TIMSS respondents are older and are expected to be better readers than PIRLS respondents, roughly 71% of students in the TIMSS sub-sample used in the analysis completed the questionnaire in a language not frequently used at home. We might, therefore, expect low literacy to plague the validity of the TIMSS background questionnaire items as well.

One way to test whether high scores on the student confidence index are simply reflective of higher literacy is to examine response patterns in the questionnaire items that comprise this scale. Specifically, it is instructive to compare student responses to positively and negatively worded items, given that negatively worded items are more difficult to answer and, therefore, require higher literacy. The distributions of scores on the positively and negatively worded items should look roughly similar if higher literacy is not driving higher scores on the confidence indices. For example, if the positively worded statement "Reading is easy for me" captures the same latent construct as the negatively worded statement "I am just not good at reading", then the distributions of scales obtained by combining positively and negatively worded statements, respectively, should be roughly similar. These distributions are plotted in Figure 8 below.

It is clear from the figure that there is evidence of a systematic difference in the distributions of positively and negatively worded items comprising the confidence index in PIRLS: while the distribution of negatively worded items for non-exceptional performers lies clearly to the left of that for exceptional performers[14], the distributions of positively worded items are roughly similar for these two groups. This evidence suggests that higher literacy among exceptional performers may be driving some of the association between confidence in reading and the probability of exceptional performance, and is a major limitation of the results presented in Section 2.7.

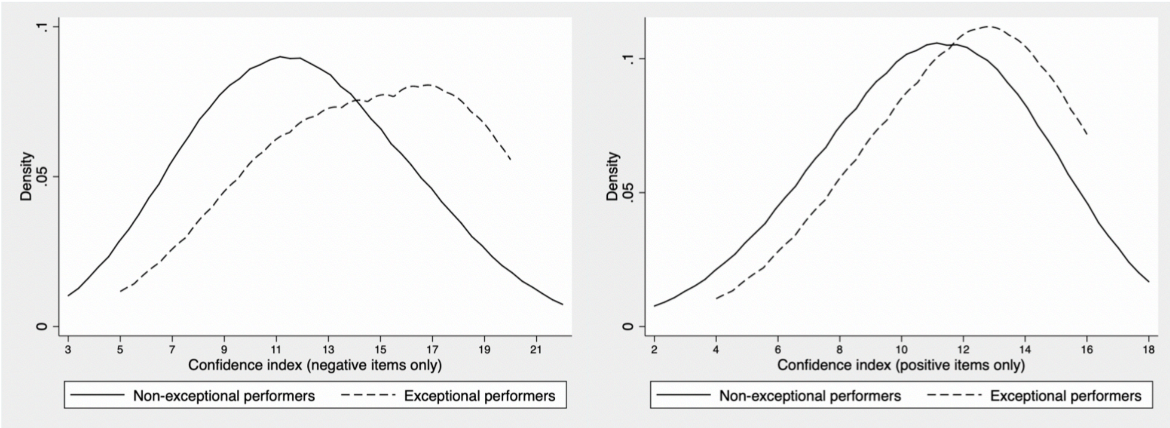
Figure 8: Distributions of negatively and positively worded items in the confidence index: PIRLS 2016



In light of the above, it is instructive to evaluate response patterns in the confidence index in TIMSS, given that these students are older and, therefore, low literacy is likely to be less of a concern than it is for the PIRLS sub-sample. Figure 9 shows the distributions of the negatively and positively worded items comprising the confidence in mathematics index in TIMSS. The figure shows much more similar distributions for the negatively and positively worded items comprising the confidence index, respectively. This provides some evidence that higher literacy among exceptional performers in TIMSS is less likely to drive the large association between confidence and the probability of exceptional performance than is the case for exceptional performers in PIRLS.

[14] Since the negatively worded items are reverse-coded, this indicates that exceptional performers systematically scored higher on these items than the comparison group.

Figure 9: Distributions of negatively and positively worded items in the confidence index: TIMSS 2015



While the different distributions for positively and negatively worded items on the confidence in reading index in PIRLS provides cause for concern, it may also be that responses to the items comprising this index are true reflections of the latent construct they attempt to capture. Another way to test whether the large coefficient on the confidence index in PIRLS is driven by illiteracy among non-exceptional performers is to limit the comparison group to literate students and estimate the same logistic regression. When limiting the sample to students who reached the low international benchmark (a level at which students are able to retrieve explicitly stated information and make straightforward inferences when reading less difficult texts (Mullis *et al.*, 2017)), the coefficient on the confidence index decreases in size but remains positive and maintains its significance. This provides some evidence that the association between confidence and exceptional performance observed in Table 3 are not entirely driven by higher literacy among exceptional performers. The fact that there is an equally large and significant association between confidence and exceptional performance in TIMSS, and little evidence of literacy differences between exceptional performers and the comparison group in the TIMSS data, provides more support for the notion that the large coefficient on student confidence in PIRLS is not entirely driven by literacy differentials between exceptional performers and the comparison group.

Another way of attempting to separate out the effect of literacy on the coefficients on confidence indices is to construct a variable that measures the difference between the negatively worded and positively worded items that comprise these indices. More literate students would have a smaller gap between scores on the negatively versus positively worded items, thus this variable could approximate literacy ability. Controlling for this variable in the logistic regressions of exceptional performance could therefore allow one to more precisely estimate the association between student attitudes and the probability of being an exceptional performer. The results of controlling for this variable (referred to as "Quality of answers") in the logistic regressions of exceptional performance are reported in Table 7 below.

Two noteworthy results emerge from attempting to parse out students' confidence and literacy ability in this way. Firstly, the coefficients on the confidence index in TIMSS remains virtually unchanged, while this coefficient decreases slightly in the PIRLS estimation, from 2.131 in Model (2) (Table 3) to 1.811. This result provides further evidence that the large and significant association between the confidence indices and the probability of exceptional performance found in the main estimation (Table 3) are not driven entirely by differences in literacy levels. Secondly, it is interesting to note that the coefficient on the student engagement index in the PIRLS estimation lose their significance entirely when controlling for literacy ability in this way. This is a noteworthy result, as it suggests that the significance of the association between student engagement and the probability of exceptional performance reported in Table 3 is driven largely by differences in literacy levels.

Table 7: Coefficients on student attitudes, controlling for difference between positively and negatively worded items (Odds Ratios)

	(1) PIRLS	(2) TIMSS
School safety and attitude indices		
Confidence in reading (PIRLS) / mathematics (TIMSS)	1.811***	2.298***
	(5.56)	(6.12)
Engagement in reading (PIRLS) / mathematics (TIMSS) lessons	1.084	0.824
	(0.54)	(1.55)
Enjoyment of reading (PIRLS) / mathematics (TIMSS)	1.192	1.020
	(1.34)	(0.19)
Sense of school belonging	1.117	0.919
	(1.24)	(0.47)
Frequency of experiencing bullying	0.806***	0.821**
	(2.84)	(2.18)
Quality of answers	0.944**	0.959*
	(2.42)	(1.69)
Observations	9,609	9,316

Notes: Models include controls for individual and school characteristics, but are not reported here. Coefficients expressed in odds ratios. Standard errors are in parentheses and clustered at the school level. Asterisks indicate statistical significance levels at *p<0.1; **p<0.05; ***p<0.01.

2.10. DISCUSSION

Despite these measurement concerns related to student attitudes, a number of important insights emerge from the results presented here. Firstly, the descriptive results provide evidence of exceptional academic performance, even in schools that achieved very poor average results in PIRLS. Over half of the exceptional performers identified in the South African PIRLS data attended schools where the median student did not reach the low international reading benchmark. This suggests it is possible for a small minority of students to achieve good academic outcomes, even in contexts of very poor average performance.

Secondly, a number of interesting results emerge from the multivariate analysis. Notably, comparing the coefficient sizes of the different student attitude indices indicates confidence in reading and mathematics is the strongest predictor of exceptional performance in these subjects, respectively. This result adds to existing evidence of a strong relationship between self-confidence in a subject and achievement in that subject, and suggests this relationship holds when studying academic achievement that exceeds expectations among socio-economically disadvantaged students. The result that student confidence is the strongest predictor of academic resilience – even when controlling for a host of factors at the level of the home and school – suggests that subject-specific self-confidence may be an important determinant of good academic results in high-poverty contexts. However, due to the limitations of the analysis presented in this chapter discussed in Section 2.9, more research will have to be conducted on the role of subject-specific self-confidence in determining learning outcomes. An important topic for future research is attempting to establish the direction of causality between student attitudes such self-confidence and exceptional academic performance.

The multivariate estimation results further indicate that girls are much more likely to be exceptional performers in reading in Grade 4. This result echoes findings from Cappella and Weinstein (2001), Vera, Valenzuela and Sotomayor (2015) and Wills and Hofmeyr (2019) who find that disadvantaged girls are more likely to be exceptional performers in reading. The lack of a gender effect in predicting exceptional performance in TIMSS is also a noteworthy result, given that international studies often find a female *disadvantage* when predicting exceptional mathematics performance in the high school grades (Cheung, 2016; Agasisti *et al.*, 2018). Furthermore, is noteworthy that the frequency with which the test language is spoken at home significantly predicts academic resilience in TIMSS; however, no such association exists in PIRLS. This suggests familiarity with English or Afrikaans may be advantageous for mathematics performance of socio-economically disadvantaged Grade 9 students in South Africa.

2.11. CONCLUSION

In this study, exceptional performers were identified in PIRLS 2016 and TIMSS 2015 data for South Africa. Three key findings emerged. First, exceptional performers could be found even in schools with very low average levels of performance. Second, exceptional performers differ systematically from their peers along a number of dimensions. Third, students' confidence in reading and mathematics is strongly associated with the probability of exceptional performance in these subjects. While we cannot infer causality from these associations, the analysis presented here is an important first step in understanding the potential role of non-cognitive skills such as student attitudes in assisting students to overcome the risks to their academic success that result not only from their socio-economically disadvantaged home backgrounds, but also the low-quality schools they attend. The result that student confidence is the strongest predictor of academic resilience, even after controlling for a host of factors at the home and school level, indicates that investigating the role of subject-specific self-confidence in determining exceptional academic performance, in particular, is likely to be a fruitful avenue for future research aimed at understanding what makes some students perform above expectations, with the view to designing policy that enables more students to do the same.

CHAPTER 3: PERSEVERANCE, PASSION AND POVERTY: EXAMINING THE ASSOCIATION BETWEEN GRIT AND READING ACHIEVEMENT IN HIGH-POVERTY SCHOOLS

3.1. INTRODUCTION

The previous chapter investigated whether non-cognitive skills such as self-confidence may assist students in overcoming the risks to their academic success that result from both their disadvantaged home backgrounds and the low-quality schools they attend. This question was approached from a resilience framework, whereby I aimed to identify the characteristics of students who manage to "beat the odds" and achieve good academic results despite their disadvantaged home and schooling backgrounds. This chapter adds to the evidence base of the association between non-cognitive skills and student achievement in South Africa by investigating whether the non-cognitive skill of grit is associated with student achievement not just for outlier students, but for the entire sample of students from township and rural schools in South Africa. Specifically, the analysis in this chapter is aimed at investigating two research questions, namely

(1) What relationship, if any, exists between grit and student achievement among this sample of students attending high-poverty schools in South Africa? and

(2) Does school quality moderate the association between grit and student achievement?

Grit is a relatively new construct in psychology which has received much attention, both in public discourse and among education researchers. Duckworth and colleagues (2007) presented the construct of "grit" as a personality trait that is highly predictive of academic performance (Poropat, 2009). Grit is defined as "perseverance and passion for long-term goals" and "entails working strenuously toward challenges, maintaining effort and interest over years despite failure, adversity, and plateaus in progress" (Duckworth *et al.*, 2007: 1087-1088). Perhaps due to the intuitive appeal of the idea that academic performance results from the combination of hard work and maintained interest, especially in the face of adversity, there has been much public debate around the idea that fostering grit could be an effective strategy for raising learning outcomes (Tough, 2011, 2016; Perkins-Gough, 2013; McKenzie, 2016; Ris, 2016; Huang and Zhu, 2017). The fact that the US Department of Education currently recommendeds that grit be taught in schools (Credé, 2018) illustrates the significant degree to which education policy makers have taken to the idea that grit could raise learning outcomes in schools.

Despite this widespread enthusiasm for the construct of grit among education researchers and practitioners alike, a number of authors have pointed out that the existing evidence of the relationship between grit and student achievement is subject to serious limitations. This chapter attempts to address two limitations of the existing grit literature that are particularly important for improving our understanding of the power of grit to predict student achievement in low- and middle-income countries (LMICs), namely (1) a dearth of evidence of this relationship from LMICs, and (2) the question of whether other educational inputs – such as school quality – moderates the association between grit and student achievement.

While there are many educational inputs that could be studied in relation to the second research question (such as IQ, family background, etc.), I focus specifically on school quality as a potential moderator of the association between grit and student achievement. This decision was informed by two main considerations. Firstly, no existing studies have investigated whether school quality interacts with grit to produce learning outcomes. Investigating such potential interaction effects thus makes a unique contribution to the existing grit literature in that it may help us to better understand the relationship between grit and student achievement. Secondly, from an education policy perspective, poor schooling quality has been identified as one of the most important factors depressing learning outcomes in poor- and middle-income countries (World Bank, 2018). If there is indeed an interdependence between grit and school quality in predicting school outcomes, this interdependence should be considered when deciding, for example, whether to focus on fostering grit as a policy measure to raise learning outcomes in contexts characterised by poor-quality schools.

I employ a number of econometric strategies to investigate these questions. Firstly, ordinary least squares (OLS) regression analysis is used to investigate the first research question – that is, whether grit is associated with student achievement among a sample of students in high-poverty schools in South Africa. I interrogate this association further by splitting the sample of schools into three terciles of school functionality and running separate OLS regressions on each school tercile. The aim of doing so is to determine whether the strength of the association between grit and student achievement varies by school functionality. Potential moderating effects of school quality on the association between grit and student achievement are investigated further by testing for interaction effects between these variables.

The analysis presented in this paper makes two important contributions to the literature on the power of grit to predict learning outcomes. Firstly, it is one of only a handful of studies that examines the association between grit and learning outcomes among schoolchildren in a middle-income country, and the first to do so in Africa. I find that the perseverance subscale of grit is highly predictive of student achievement when considering the sample of township and rural schools in its entirety, an association that remains significant across all terciles of school functionality when splitting the sample into terciles and estimating this relationship separately for each tercile. This result makes an important contribution to the grit literature in that (1) it introduces evidence from high-poverty schools in an African country to the international evidence base of the association between grit and student achievement, and (2) it provides evidence that this aggregate association is not driven by a strong association in a few relatively more functional schools. That is, the perseverance subscale of grit has a strong positive relationship with student achievement, even among students in poorly functioning schools.

Secondly, in terms of the second research question, I present evidence of moderating effects between grit and school quality in predicting learning outcomes. Specifically, the results provide evidence of significant interaction effects, suggesting that the perseverance subscale of grit, specifically, interacts meaningfully with school functionality to predict reading test scores. Secondly, further interrogation of these interaction effects produces evidence that the nature of the interaction between perseverance and school functionality is not uniform across the distributions of these variables. This result makes an important contribution to the literature on interaction effects between non-cognitive skills and other education inputs, since it provides empirical evidence for the theoretical possibility that the nature of interaction effects between non-cognitive skills and other educational inputs may vary at different points of the distributions of both non-cognitive skills and the inputs they interact with.

The chapter is organised as follows: Section 3.2. presents an overview of the literature on the association between grit and student achievement, focussing specifically on studies that study this association in LMICs. The conceptual framework informing the analysis is also set out in Section 3.2. Details of the estimation sample used in this study and key measures used in the econometric analysis are presented in Section 4.3. Before proceeding to the multivariate analysis, Section 3.4. considers key issues related to the measurement of grit among this sample of students. The multivariate estimation results are presented in Section 3.5. These results – as well as their limitations – are discussed in Section 3.6., and Section 3.7. concludes.

3.2. LITERATURE REVIEW

A comprehensive account of the existing evidence regarding the association between grit and student achievement from high-income countries is provided in Credé, Tynan and Harms' (2017) meta-analysis of the grit literature. As such, instead of repeating this evidence here, I limit my review of the grit literature to studies that investigate the association between grit and student achievement in LMICs. To my knowledge, there are only four peer-reviewed studies that investigate this question.

Tovar García (2017) compares the achievement of migrant versus native Grade 9 students in Russia, and, finding no significant differences in achievement between migrant and native students despite the relative home background disadvantage of migrant students, investigates whether grit plays a role in explaining why the usual relationship between SES and student achievement is not observed among these students. He finds that grit has significant effects in explaining in explaining this phenomenon, concluding that grit may be an important factor that enables migrant students to achieve academic outcomes on par with their native counterparts.

In the first study to examine the association between grit and student achievement in Africa, Mason (2018) found that grit was significantly associated with achievement among a sample of 121 undergraduate students at a South African university. Although instructive in suggesting that grit may be an important predictor of achievement in the African context, the generalisability of Mason's results is limited by the small and selective nature of his sample, as well as the fact that he does not control for any other factors that might impact on achievement in his estimation. The analysis in this chapter therefore builds on this evidence by considering a larger sample of students, and including a number of controls at the level of the individual, home, and school, in the analysis in an attempt to better isolate the association between grit and achievement. The present study also differs from Mason's in that it considers the association between grit and achievement among primary school students.

My own work with Gabrielle Wills (2019) was the first to investigate the association between grit and student achievement among schoolchildren in the African context. Using the same resilience framework as employed in the previous chapter of this thesis, our study investigated whether the "perseverance of effort" subscale of grit predicted the probability of being identified as an exceptional performer among students in high-poverty contexts. Our results, based on the same dataset of high-poverty schools in South Africa that is used in this chapter, showed that the perseverance subscale of grit was the strongest predictor of being a positive outlier in reading performance of all measured household, classroom, and school factors. This evidence is suggestive of a strong association between the perseverance subscale of grit and student achievement, even in high-poverty schools in a middle-income country.

In the most recent contribution to the evidence base of the association between grit and student achievement from LMICs, He *et al.* (2021) examine this association among a sample of 2,931 Grade 7 students in rural China. Their study investigates the important question of whether cognitive ability moderates the association between grit and achievement in their sample – that is, whether the strength of the association between grit and achievement varies by cognitive ability. Although Light and Nencka (2019) also investigate potential moderating effects between grit and IQ in predicting student achievement, He *et al.*'s (2021) study makes an important contribution to the existing evidence base of the association between grit and achievement in that it is the first to consider the question of potential moderating effects between grit and IQ using evidence from students in high-poverty rural areas in a middle-income country. Their main result – that grit is not associated with achievement among low-IQ students – makes an important contribution to the existing evidence base of the association between grit and achievement in that it is the first evidence to suggest that this association does not hold across the ability distribution.

The analysis in this chapter is aimed at adding to this evidence base of the relationship between grit and student achievement in LMICs. In particular, the analysis builds on Wills and Hofmeyr's (2019) study by estimating the association between grit and student achievement for the full Leadership for Literacy dataset. That is, while Wills and Hofmeyr test whether grit is associated with the probability that a student will be a positive outlider in reading outcomes, the analysis in this chapter investigates whether a positive relationship between grit and student achievement is observable across the full distribution of reading scores for the Leadership for Literacy sample. The analysis further adds to the evidence base of the relationship between grit and student achievement by considering whether school functionality moderates the strength of this association.

Moderating effects between grit and other educational inputs

To my knowledge, this study is the first to investigate potential moderating effects between grit and school quality in predicting student achievement. A number of studies do however investigate potential moderating effects between non-cognitive skills – such as conscientiousness and motivation – and other inputs in producing educational outcomes (Light and Nencka, 2019). Following Ross and Mirowsky (2006), these studies often employ a framework of comparative advantage, whereby non-cognitive skills are conceived of as inputs (endowments) in the production of educational outcomes, along with other inputs, such as cognitive ability and family background. In their investigation of potential substitution effects between grit and cognitive ability, Light and Nencka (2019) explain the logic behind this reasoning as follows:

> "Students are assumed to behave optimally (given their current information) in deciding how to combine inputs, and students with a relative paucity of cognitive ability must choose between intensifying their use of other inputs, including grit, or failing to achieve… Whether substituting conscientiousness and agreeableness for family resources or substituting grit for cognitive ability, the hypothesis is that students who lack a given input exploit their comparative advantage as a means of achieving success."

According to this hypothesis (referred to as the "resource substitution" hypothesis in the literature (Damian *et al.*, 2015)), then, students who have less of a given educational input (for example, school quality) are expected to experience higher "returns" to non-cognitive skills when predicting achievement than their counterparts who have more of that input. Damian *et al.* (2015) put forth two other potential hypotheses that could describe the nature of the interaction between non-cognitive skills and other educational inputs, namely the "Matthew effect" hypothesis and the "independent effects" hypothesis. The Matthew effect, or "rich get richer" hypothesis, predicts the opposite of the resource substitution hypothesis, that is, that students who have relatively *more* of a given educational input will experience higher returns to non-cognitive skills. That is, if we consider potential interaction effects between grit and school quality in predicting student achievement, the Matthew effect hypothesis posits that the effect of being "gritty" would be greater for students who have more access to learning resources, as they would be able to use these resources to achieve better results. Students who attend lower-quality schools will have fewer resources at their disposal to achieve academically, and so they would exhibit lower returns to being gritty. Lastly, according to the independent effects hypothesis, the relationship between non-cognitive skills and student achievement will not vary across students who have different endowments of a given educational input. Thus, the independent effects hypothesis predicts that there will be no interaction effects between non-cognitive skills and a given educational input in predicting student achievement.

The analysis in this chapter therefore seeks to determine whether there is evidence of moderating effects between school quality and grit in predicting student achievement among a sample of township and rural schools in South Africa, and, if so, which of the above hypotheses (the resource substitution, Matthew effect, or independent effects hypothesis) is supported by the data. The data used in this study is particularly suited to investigating this question due to the contexts of severe socio-economic deprivation that characterise the schools in the sample. In addition, the contexts that characterise the schools used in this sample are similar to the contexts that characterise many LMICs. In this sense, estimating potential moderating effects between school quality and grit using this sample of students may provide important insights regarding these effects for students who face similar socio-economic disadvantage in other parts of the world. For example, from an education policy perspective, it is important to understand potential moderating effects between school quality and grit, since this evidence would have implications for policymakers deciding whether to focus on fostering grit as a policy measure to raise learning outcomes in contexts characterised by poor-quality schools.

3.3. DATA AND METHODOLOGY

3.3.1. DATA

The data used was gathered for a project entitled “Leadership for Literacy”, which was aimed at understanding resilience and exceptionalism in high-functioning township and rural primary schools in South Africa. In 2017, literacy tests were administered to over 2600 Grade 6 students in 60 primary schools. The Leadership for Literacy project was aimed specifically at identifying exceptional schools matched to underperforming pairs in challenging contexts, so the sample consists only of schools located in township or rural areas (for further discussion of the sampling process see Taylor, Wills, and Hoadley (2019)). The sample is representative neither at the national nor provincial level. Nevertheless, due to the matched schools design on which the sampling process was based, and the wide geographic dispersion of these schools within provinces, they are a close representation of schools in socio-economically disadvantaged areas in three provinces of South Africa (KwaZulu-Natal, Limpopo and Gauteng).

Most quantitative studies of education in South Africa use assessment data from only one point in time. However, there is consensus in the international literature that such cross-sectional assessment data are subject to considerable measurement error (Agostinelli, Saharkhiz and Wiswall, 2019). A key advantage of the Leadership for Literacy data is that the reading and literacy assessments were administered twice, at the beginning of the school year and again towards the end of the same school year. The pre- and post-test have high levels of reliability. The longitudinal dimension of the Leadership for Literacy assessment data offers a unique opportunity to reduce measurement error in reading scores which most quantitative studies of education in South Africa have not been able to do. The assessment consisted of a silent reading comprehension test that was administered to an entire class of Grade 6 students in each school. Of the original pre-test sample of 2 656 students, 2 383 wrote the post-test, indicating a low attrition rate of 11%. The two comprehension tests consisted of released items from previous rounds of the grade 4 PIRLS assessment. Permission was received from the International Association for the Evaluation of Educational Achievement (IEA) for their use.

The tests were conducted in English and, in this respect, reading achievement in this study is largely defined in terms of reading achievement in English. This may be criticised as a measure of overall reading achievement since students may perform badly in English reading but well in other subjects or languages. However, English language proficiency is a necessary condition for academic success in South African schools. Although schools are at liberty to choose one of the 11 official languages as their medium of instruction in foundation phase grades (Grades R-3)[15], all are required to teach in English or Afrikaans from Grade 4 onwards. If children cannot read and write in English by the end of Grade 3, it is very difficult for them to access the curriculum.

In addition to the comprehension tests, the research team also administered student background questionnaires and collected information about school characteristics through fieldworker observations. Information from both the student background questionnaires and fieldworker observations was used to derive the individual, home and school variables included in the multivariate analysis.

3.3.2. MEASURES

In the main model, reading performance is measured as the mean silent comprehension test score between the pre-test and the post-test. Students could score a maximum of 32 points in the silent comprehension test. Raw comprehension scores were standardized for ease of interpretation. The measure of school quality included in the multivariate analysis is derived from fieldworker observations. This measure includes information about school infrastructure, learning materials, and instructional time. Principal components analysis was used to derive an index of school functionality based on 10 of these measures. The index has a high scale reliability coefficient of 0.70[16]. This index was used to split the sample into terciles of school functionality, each containing 20 schools. Due to larger class sizes in the bottom tercile, there are more students in the bottom tercile (868 students) compared to the second and third terciles (743 and 772 students in each tercile, respectively).

[15] Grade R is the “reception” year in South Africa, that is, the grade preceding Grade 1, equivalent to Kindergarten in the United States.

[16] A graph of the Eigenvalues of the PCA used to construct the school functionality index can be found in Appendix B (Figure B1).

Table 8 shows differences between the three groups of schools, in terms of the variables included in the school functionality index at the school and classroom level, respectively. The proportions shown are calculated at the student level. The figures show that even though most of the schools in the sample are no-fee schools[17], there is significant variation in functionality between the three groups of schools. For example, while less than 10 percent of students in the bottom tercile attended schools where most toilets worked, this proportion was 85 percent for students in the top functionality tercile. Differences in the proportions of students in classes that exceeded 50 students (the variable labelled "large class" in Table 8) is particularly striking, with almost half (47 percent) of students in the bottom tercile of schools being in such large classes, and none of the Grade 6 classes in the top tercile exceeding 50 students. It should be noted that large classes can be considered a measure of school functionality, and not simply school resources, in that they indicate how teachers are utilised in terms of timetabling within schools. 55 out of the 60 schools in the sample are no-fee schools[18] and would, therefore, not have been able to appoint extra teachers with school governing body funds. Since teacher allocations are based on enrolments, the number of teachers in each school should – at least theoretically – be proportional to the number of students. It is in this sense, then, that differences in class sizes reflect differences in the effectiveness with which school resources – in this case, teachers – are utilised. The effects of such large classes are reflected in the variables indicating whether the Grade 6 classrooms were cramped, whether there were enough chairs or desks for all students, and whether there were enough textbooks for all students, with these proportions decreasing with school functionality tercile.

Table 8: School and classroom characteristics by functionality tercile

	Tercile 1	Tercile 2	Tercile 3
Library	22%	62%	89%
Most toilets work	7%	56%	85%
Teachers missing	29%	41%	34%
School-wide reading period	17%	26%	51%
Large class	47%	21%	0%
Classroom cramped	42%	24%	12%
Not enough chairs	99%	40%	11%
Not enough desks	95%	44%	12%
Desks broken	59%	24%	6%
Enough textbooks	32%	55%	61%

Notes: "Teachers missing" indicates whether at a certain point during the school day, fieldworkers observed any classes where students were present but there was no teacher in the classroom. "School-wide reading period" indicates whether the school timetable included a daily period dedicated to reading. "Large class" indicates whether there were more than 50 students in the Grade 6 class participating in the study (derived from the number of students who wrote the literacy assessment in each Grade 6 class).

[17] No-fee schools make up roughly the poorest 60% of schools in South Africa (referred to as Quintile 1-3 schools), which are typically under-resourced and characterized by legacies of dysfunction (Wills, 2017). School quintiles were originally constructed using Census information on the infrastructural development of the surrounding area to inform student funding allocations in a pro-poor manner. These schools are technically not allowed to charge fees due to the relative poverty of the communities they serve and the South African Constitution's commitment to providing free basic education.

[18] Five of the schools are low-fee schools with fees under than R2000 per annum.

Student wealth was measured using information about 11 assets students indicated having in their homes in the student background questionnaire. The proportions of students who indicated they have certain assets in their homes is shown in Table 9. It is clear from the figure that although the sample of schools was purposefully chosen to represent students from low socio-economic backgrounds, there is significant variation in material resources in the homes of the students in the Leadership for Literacy sample. It is further evident from Table 9 that this variation is highly correlated with school functionality, with students in the bottom school tercile facing more resource deprivation at home, on average. For example, while only 11 percent of students in the bottom tercile of schools indicated that they had running water in their homes, this proportion was 39 percent for students in the top school tercile. Similarly, only 24 percent of tercile 1 students indicated having a flush toilet in their homes, whereas 59 percent of tercile 3 students did so. In addition to significant variation in the material wealth of students in the Leadership for Literacy sample by school functionality, Table 9 also points to the severity of resource deprivation faced by the students in this sample overall, providing further motivation for using this sample to study the association between grit and achievement in high-poverty contexts. Principal components analysis was used to derive an index of these 11 assets, which is used as the measure of student SES in the multivariate analysis. This index also has a high scale reliability coefficient, at 0.79[19]. While it is standard practice in the literature to include information on parental education in measures of student SES, unfortunately information on parental education was not collected in the Leadership for Literacy student background questionnaire. The measure of student SES used in the analysis that follows is, therefore, strictly a measure of material wealth. The SES of a school's student body was taken as the mode of this asset index score for the Grade 6 pupils in that school.

Table 9: Home assets by school functionality tercile

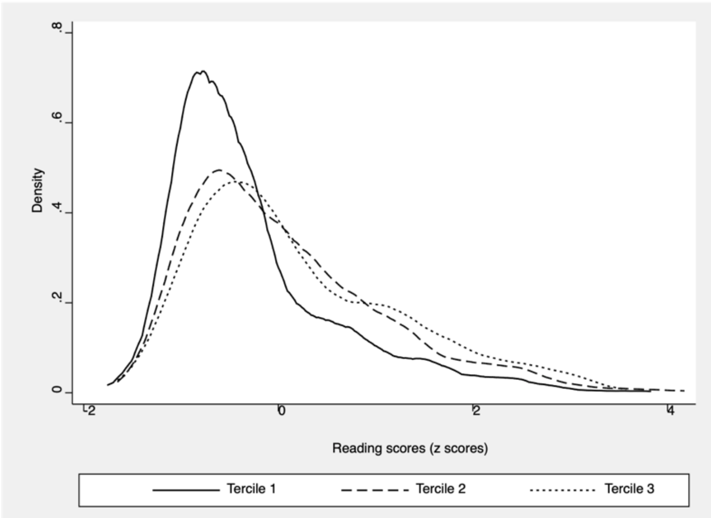
	Tercile 1	Tercile 2	Tercile 3
Cell phone	94%	97%	97%
Fridge	86%	92%	96%
TV	86%	93%	98%
Microwave	55%	69%	80%
Car	52%	59%	64%
Running water	39%	71%	81%
Running water (hot)	11%	26%	35%
Internet	40%	43%	53%
Computer	35%	45%	50%
Washing machine	28%	54%	61%
Toilet	24%	54%	59%

Notes: Information about home assets was collected in the student background questionnaire. Students had to tick which assets they had in their home from a list. Pictures were included to minimize the potential effect of low literacy on the accuracy with which students responded to these questionnaire items.

[19] A graph of the Eigenvalues of the PCA used to construct the student asset index can be found in Appendix B (Figure B2).

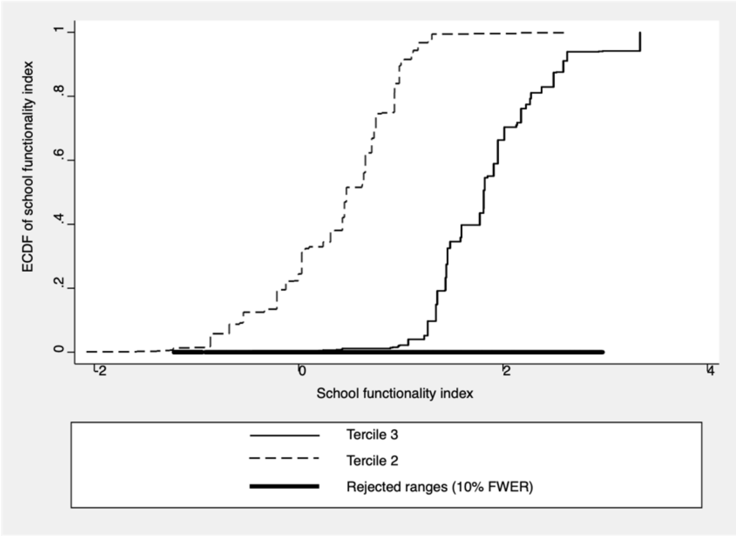
Figure 10 shows the distribution of reading scores for each tercile of school functionality. As expected, the distribution of comprehension test scores of students in the bottom school tercile lies clearly to the left of the other two terciles of school functionality. It is interesting to note, however, that there is little variation in the distributions of reading scores of students in the second and third terciles. In light of this, I test whether the middle and top school functionality terciles are statistically significantly different from each other, in terms of scores on the school functionality index, by comparing the estimated cumulative distribution functions (CDF's) of the school functionality index for students in tercile 2 and 3 schools, respectively. These distributions are plotted in Figure 11 below. The range where the null hypothesis of identical distributions is rejected is also plotted in the figure (the thick horizontal line near the bottom). It is clear from the figure that the two distributions are significantly different from each other at all points of the distributions of the school functionality index. This provides strong evidence that the tercile 2 and tercile 3 distributions are not identical. I therefore proceed with the analysis using these three terciles of school functionality.

Figure 10: Kernel density distribution of reading scores by school functionality tercile



Notes: Kernel density distribution using epanechnikov, bandwidth = 11.8405

Figure 11: CDF's of school functionality index, by school tercile



Notes: Cumulative distribution functions (step functions) of the school functionality index are plotted for school terciles 2 and 3, respectively. The thick horizontal line shows the range where CDF equality is rejected. Simulated p-value for the global test of equality of the two CDF's < 0.0001.

3.4. MEASURING GRIT

Before attempting to investigate the main research questions of this chapter, it is necessary to assess the reliability with which grit is measured in the Leadership for Literacy sample of students. Student grit is measured using responses to an adapted version of the short grit scale (Grit-S) (Duckworth and Quinn, 2009), where students had to rate themselves on eight items, choosing from “That’s not at all like me”, “That’s not really like me”, “That’s sometimes like me”, and “That’s a lot like me”. The eight questions comprising the adapted version of the short grit scale can be found in Appendix B.

As is the case for the measurement of student attitudes in PIRLS and TIMSS discussed in the previous chapter, of particular concern is that participating students in the Leadership for Literacy sample may not have the required literacy skills to answer questionnaire items meaningfully. Therefore, it is necessary to test for evidence that differential levels of literacy might be driving differential responses to the grit questionnaire items. Since reading scores increase with school functionality, I calculate the scale reliability coefficients of the grit scale and its subscales by school functionality tercile (Table 10) as a test for possible effects of reading ability on responses to the grit questionnaire. This also allowed for testing of the possibility that differences in scale reliability by school functionality drive any observed differences in the association between grit and achievement by school functionality. Table 10 shows that at 0.43, 0.46 and 0.47, the alpha coefficients of the full grit scale are not high enough for the scale to be considered internally reliable for any of the school terciles. The alpha coefficients of the two subscales of grit indicate that low internal reliability of the ‘consistency of interest’ subscale is driving the low alpha values for the full grit scale. Removing the ‘consistency of interest’ items from the grit scale, that is, limiting the scale to only the ‘perseverance of effort’ subscale, increases the alpha coefficients across all terciles of school functionality. Since all the consistency of interest items are negatively worded, this evidence adds to the evidence presented in the previous chapter of construct validity issues introduced when negatively worded items are included in student background questionnaires, especially in low literacy contexts (see for example Bofah and Hannula (2015)).

Table 10: Scale reliability coefficients for the grit scale by school functionality tercile

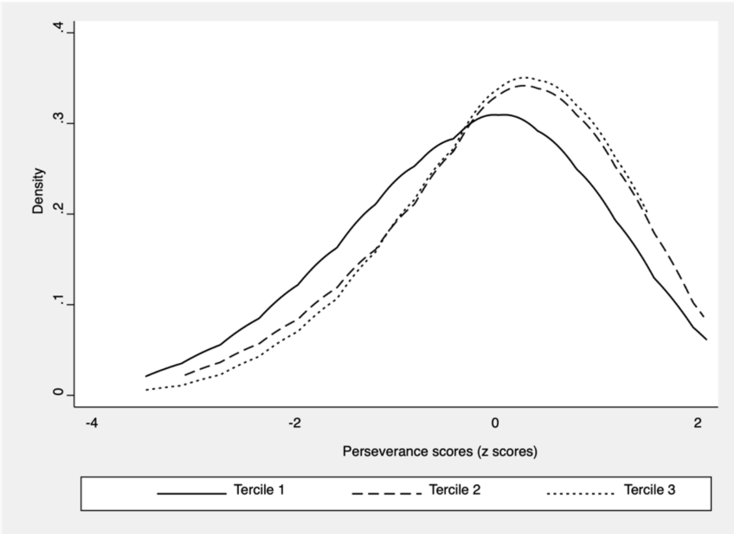
Alpha coefficients			
	Full grit scale	'Perseverance of effort' subscale	'Consistency of interest' subscale
Tercile 1	0.43	0.54	0.22
Tercile 2	0.46	0.55	0.32
Tercile 3	0.47	0.52	0.40

The results in Table 10 do not provide any evidence of differences in the scale reliability of grit and its two subscales by school functionality. The results in the table do indicate, however, that the grit scale is not measured with sufficient internal reliability among the Leadership for Literacy sample. Although the alpha coefficients of the 'perseverance of effort' subscale are considered "poor" (Kline, 2000), a number of peer-reviewed studies that investigate association between grit and academic achievement make use of grit measures with similarly low alpha coefficients for the perseverance subscale of grit (see for example Light and Nencka (2019) and Wills and Hofmeyr (2019)). For this reason, a decision was made to use only the perseverance subscale of grit in the analysis that follows.

3.5. MULTIVARIATE ESTIMATION RESULTS

As per the two main research questions posed at the beginning of this chapter, the main objectives of the multivariate estimation strategy are to determine (1) whether the perseverance subscale of grit is associated with reading outcomes for the full Leadership for Literacy sample, and (2) whether there is evidence of moderating effects between school quality and grit in reading achievement. Before examining the relationship between perseverance and reading scores in a multivariate context, however, it is instructive to investigate whether there are differences in the overall levels of perseverance by school functionality. To this end, the distributions of perseverance scores for each school tercile are shown in Figure 12. The figure shows that perseverance scores increase slightly with school functionality. However, it is interesting to note that the differences in perseverance scores by school tercile are much smaller than the differences in reading comprehension scores between school terciles (shown in Figure 10 above). The fact that there are minor differences in perseverance scores between school terciles but large differences in reading scores between school terciles is suggestive of a stronger association between perseverance and reading scores in more functional schools. This hypothesis is investigated further in a multivariate context by running separate ordinary least squares (OLS) regressions[20] for each tercile of school functionality. In order to answer the first research question, however, I first estimate the association between perseverance and reading scores for the full sample of students. These results are presented in the first column of Table 11. In addition to student SES, the following controls are included in the multivariate regressions: gender, age, the frequency of English use at home, whether the student lives with their mother and father, whether the student's parents are employed, the language of learning and teaching (LOLT) of the school in the foundation phase, and province.

Figure 12: Kernel density distributions of perseverance scores by school functionality tercile



Notes: Kernel density distribution using epanechnikov, bandwidth = 11.8405

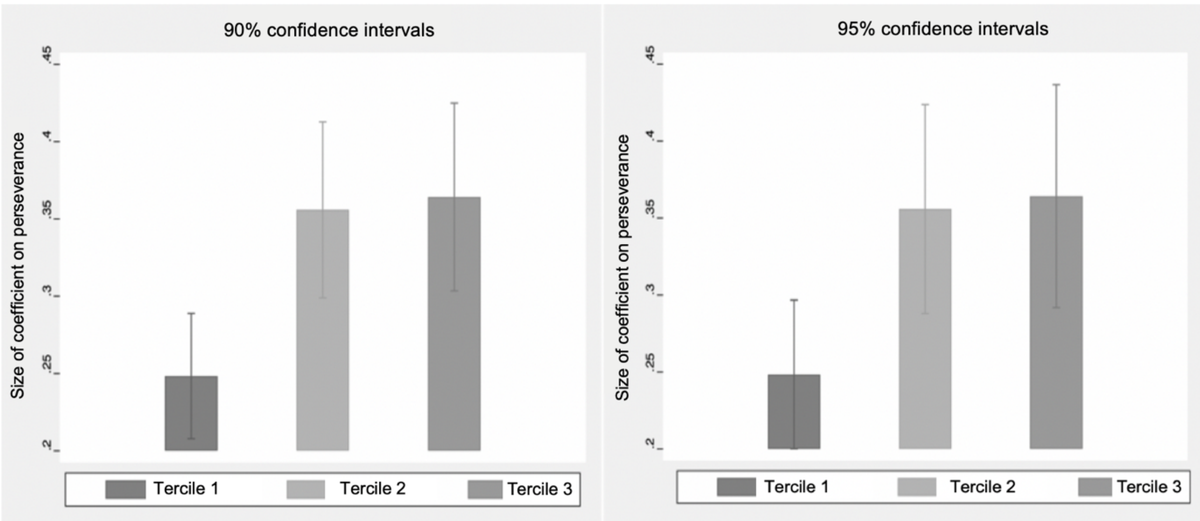
3.5.1. ESTIMATION OF MAIN EFFECTS

The results in Table 11 indicate a large and significant association between perseverance and reading comprehension scores for the full Leadership for Literacy sample, even when controlling for a host of factors at the individual, home, and school level. Importantly, this association holds for all school functionality terciles. In terms of the first research question, then, the results presented in Table 11 show that the perseverance subscale of grit exhibits a positive association with reading outcomes, even for students who attend low quality schools. This result makes an important contribution to the existing evidence base of the relationship between grit and student achievement in high-poverty contexts in LMICs, which thus far has not investigated whether the strong association between grit and student achievement observed in high income countries holds across schools with varying levels of functionality in high-poverty contexts.

[20] The association between perseverance and reading scores was also modelled using a hierarchical linear model (HLM) to account for the multilevel nature of the data, whereby students are nested within schools. Since the results from the hierarchical model are not significantly different from those obtained using OLS, a decision was made to report only the results from the OLS estimation in the main text. The results from the hierarchical model can be found in Table B1 of Appendix B.

In addition to the coefficient on the perseverance subscale of grit being positive and significant across all school terciles, the results presented in Table 11 provide some preliminary evidence of moderating effects between school functionality and perseverance in predicting reading scores, with the size of the coefficient on perseverance varying across with school functionality terciles. As such, it can be noted that a standard deviation increase in perseverance scores is associated with a quarter of a standard deviation (24.8%) increase in reading scores among students in tercile 1 schools, compared to 35.6% and 36.4% of a standard deviation increases among students in tercile 2 and 3 schools, respectively. The fact that the coefficient on perseverance increases with school functionality is suggestive of a Matthew effect in the interaction between school functionality and grit, whereby students who already have the advantage of attending more functional schools benefit more from being gritty, in terms of reading scores, than students who attend less functional schools. To test whether these differences in the strength of the association between perseverance and reading scores by school functionality are statistically significant, the coefficients on perseverance reported in Table 11 are plotted in Figure 13, with 90% and 95% confidence intervals, respectively, around these estimates. The figure shows the coefficient on perseverance in tercile 1 schools is statistically significantly different from the coefficient on perseverance in tercile 3 schools at the 90% level, but not at the 95% level. The next section investigates this result further by modelling interaction effects between perseverance and school functionality.

Figure 13: Size of coefficients on perseverance by school functionality tercile



Source: Leadership for Literacy

The regression results reported in Table 11 further point to a large gender gap in favour of girls, in terms of reading scores among the Leadership for Literacy sample. This result is consistent with the result presented in Chapter 2 that girls are more likely to be identified as exceptional performers in PIRLS. As mentioned in Chapter 2, this pro-girl gap in academic performance is well-documented in the local as well as international literature (Steinmayr and Spinath, 2008; Spaull and Makaluza, 2019), especially in reading in the primary school grades. It is interesting to note that a gender gap remains even after controlling for perseverance, as a number of authors present evidence that gender differences in socio-emotional skills may contribute to the pro-girl advantage in achievement that is documented in many countries (see for exmaple Steinmayr and Spinath, 2008; Dercon and Singh, 2013; Spinath, Eckert and Steinmayr, 2014). The fact that the gender gap in favour of girls remains after controlling for perseverance indicates that gender differences in perseverance, at least, do not explain away the pro-girl advantage in terms of reading scores among the Leadership for Literacy sample of students. The roll of non-cognitive skills in South Africa's pro-girl advantage is investigated explicitly in Chapter 4. Other covariates that are significantly associated with reading scores across school terciles include being over-age, having attended Grade R, and being from a wealthier home.

Table 11: Estimation of comprehension test scores

	Full sample	Tercile 1	Tercile 2	Tercile 3
Perseverance (z-scores)	0.306***	0.248***	0.356***	0.364***
	(0.018)	(0.037)	(0.026)	(0.048)
Female	0.256***	0.314***	0.212***	0.284***
	(0.036)	(0.060)	(0.061)	(0.082)
Over-age	-0.274***	-0.190***	-0.314***	-0.336***
	(0.048)	(0.052)	(0.064)	(0.076)
Frequency of English use at home	0.170***	0.114**	0.244***	0.101
	(0.037)	(0.063)	(0.078)	(0.075)
Attended Grade R	0.187***	0.166**	0.111	0.327***
	(0.059)	(0.067)	(0.146)	(0.117)
Asset index (z-scores)	0.156***	0.178***	0.170***	0.200***
	(0.023)	(0.041)	(0.059)	(0.052)
Asset index^2	0.057***	0.053	0.032	0.017
	(0.018)	(0.032)	(0.048)	(0.056)
Lives with mother	0.073	0.089*	0.092	-0.027
	(0.039)	(0.039)	(0.084)	(0.074)
Lives with father	0.011	0.046	0.039	-0.103
	(0.037)	(0.076)	(0.091)	(0.062)
At least one parent employed	0.136***	0.227***	0.022	-0.019
	(0.047)	(0.045)	(0.104)	(0.086)
School functionality index	0.112***	0.222	-0.004	0.051
	(0.028)	(0.124)	(0.157)	(0.167)
School uses English as the LOLT in the foundation phase	0.036	-0.072	0.074	0.068
	(0.038)	(0.115)	(0.089)	(0.095)
Peer SES	0.038***	0.020	0.029	0.356**
	(0.012)	(0.053)	(0.097)	(0.142)
Constant	-1.468***	-0.929***	-1.271***	-1.507***
	(0.134)	(0.168)	(0.245)	(0.317)
R^2	0.31	0.32	0.28	0.27
N	2,383	868	743	772

Notes: Asterisks indicate statistical significance levels at *p<0.1; **p<0.05; ***p<0.01. Standard errors (clustered at the school level) are reported in brackets. 'Over-age' indicates that the student is one or more years older than they should be in Grade 6 (12 years). 'Frequency of English use at home' is a categorical variable coded as follows: 0 "Never or almost never";1 "Sometimes"; 2 "Always or almost always". Province controls are included but not reported here. Source: Leadership for Literacy.

3.5.2. ESTIMATION OF INTERACTION EFFECTS

Variation in the strength of the association between perseverance and reading scores by school functionality was further interrogated by again running the regression reported in Table 11 on the full Leadership for Literacy sample, with interaction effects between perseverance and school functionality. The results of this regression are reported in Table 12. The same factors are controlled for as in the regressions in Table 11 but are not reported. The results in the table provide further evidence of a Matthew effect in the strength of the association between perseverance and reading scores by school functionality tercile: In Tercile 1 schools, students high in perseverance had reading scores 66.0% of a standard deviation higher than students low in perseverance, on average. Students high in perseverance in Tercile 2 and 3 schools, on the other hand, achieved reading scores 83.9% and 88.7% of a standard deviation higher than students low in perseverance in tercile 1 schools, on average. In other words, compared to students low in perseverance in Tercile 1 schools, students high in perseverance in Tercile 2 and 3 schools have a greater advantage in terms of reading scores than their counterparts in Tercile 1 schools who also scored high in perseverance. Wald tests of significance show that these differences are statistically significant at the 95% level.

Table 12: Interaction effects between perseverance and school functionality

	Low perseverance	Medium perseverance	High perseverance
Tercile 1	Omitted	0.204*** (0.058)	0.660*** (0.103)
Tercile 2	0.080 (0.106)	0.567*** (0.112)	0.839*** (0.123)
Tercile 3	0.147 (0.146)	0.465*** (0.133)	0.887*** (0.131)

Notes: Asterisks indicate statistical significance levels at *p<0.1; **p<0.05; ***p<0.01. Standard errors (clustered at the school level) are reported in brackets. The model includes the following controls: Gender, age, frequency of English use at home, whether the student attended Grade R, student SES, whether the student lives with either biological parent, whether the student has at least one employed parent, the school's LOLT in the foundation phase, peer SES, and province.

To further test the significance of the interaction effects reported in Table 12, I test for the marginal effect of being in a different school tercile on the association between perseverance and reading scores. To do this, I estimate the predictive margins of perseverance for each school tercile, and compare these predictive margins across school terciles. These results are reported in Table 13. The estimates reported in the table suggest that the marginal effects of school functionality tercile on the association between perseverance and reading scores are only significant between Terciles 1 and 2. Specifically, moving from "low" to "medium" on perseverance in Tercile 2 schools is associated with a larger increase in reading scores than is the case for students in Tercile 1 schools (indicated by the positive and significant predictive margin), whereas moving from "medium" to "high" on perseverance is associated with a smaller increase in reading scores among Tercile 2 students than Tercile 1 students (indicated by the negative and significant predictive margin). This is an interesting result, since it suggests that although the marginal effect of being in a Tercile 2 versus Tercile 1 school on the association between perseverance and reading scores is statistically significant, the direction of this marginal effect is not uniform across different levels of perseverance. That is, while Tercile 2 students who scored "medium" compared to "low" on perseverance are at an added advantage in terms of reading scores over their Tercile 1 counterparts, the opposite is true for Tercile 2 versus Tercile 1 students who scored "high" compared to "medium" on perseverance.

Table 13: Contrasts of interaction effects sizes between perseverance and school functionality tercile

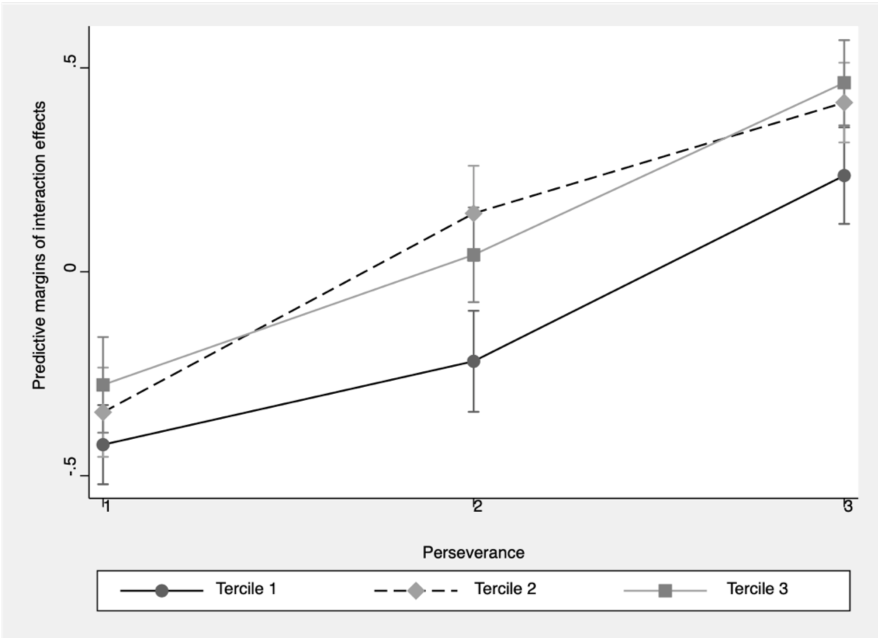
	Contrast	P>F
Perseverance#School_tercile		
2. Medium vs 1. Low (2 vs 1)	0.283*** (0.107)	0.008
2. Medium vs 1. Low (3 vs 2)	-0.169 (0.112)	0.132
3. High vs 2. Medium (2 vs 1)	-0.184** (0.108)	0.090
3. High vs 2. Medium (3 vs 2)	0.150 (0.105)	0.152
Joint		0.112

Notes: Asterisks indicate statistical significance levels at *p<0.1; **p<0.05; ***p<0.01. Standard errors (clustered at the school level) are reported in brackets. The model includes the following controls: Gender, age, frequency of English use at home, whether the student attended Grade R, student SES, whether the student lives with either biological parent, whether the student has at least one employed parent, the school's LOLT in the foundation phase, peer SES, and province.

These results are presented graphically in Figure 14. It is clear from the figure that moving from "low" to "medium" on perseverance has the highest return (in terms of reading scores) for students in Tercile 2 schools (the dotted line in the figure), however this marginal effect decreases when moving from "medium" to "high" perseverance. The figure further shows that although the line representing the predictive margins of the interaction between perseverance and functionality for students in Tercile 3 schools (the light grey solid line) lies above that for students in Tercile 1 schools (the dark grey solid line), these lines are roughly parallel, suggesting the marginal effect of perseverance on reading scores is similar for students in Tercile 3 versus Tercile 1 schools. Therefore, while the results from the main estimation (Table 11 and Figure 13) provide some support for the Matthew Effect hypothesis across all three terciles of school functionality, the results in Table 13 and Figure 14 suggest this effect is only observable when moving from "low" to "medium" in perseverance, and only when comparing students in Tercile 2 versus Tercile 1 schools. By contrast, the difference in reading scores between students who scored "medium" versus "high" in perseverance was *smaller* among students in Tercile 2 schools compared to students in Tercile 1 schools. The latter result provides evidence of a resource substitution hypothesis (a higher return to perseverance among students in *less* functional schools), rather than a Matthew effect hypothesis.

While these results may seem to contradict each other, it is important to note, as Light and Nencka (2019) do, that it is possible for Matthew effects, resource substitution effects, and independent effects to exist side-by-side. That is, it is possible that the direction and significance of interaction effects between two inputs varies depending on where on the distributions of both inputs one tests for these effects. The evidence presented in Table 13 and Figure 14 suggests that this is the case for the nature of the interdependence between perseverance and school functionality for this sample of students. This result makes an important contribution to the literature on interaction effects between non-cognitive skills and other education inputs, since it provides empirical evidence for the theoretical possibility that the nature of interaction effects between non-cognitive skills and other educational inputs may vary at different points of the distributions of both non-cognitive skills and the inputs they interact with.

Figure 14: Margins plot of interaction effects between perseverance and school functionality



Notes: Predictive margins of interacting perseverance with school functionality tercile are plotted on the y-axis. Perseverance scores (plotted on the x-axis) coded as 1 "Low", 2 "Medium", and 3 "High". Error bars indicate 95% confidence intervals. Source: Leadership for Literacy.

3.6. DISCUSSION

This chapter investigated the association between the perseverance subscale of grit and reading achievement among a sample of Grade 6 students in township and rural schools in South Africa. The analysis undertaken in this chapter makes an important contribution to the literature in that it is one of few studies that estimates the relationship between grit and academic achievement in a middle-income country, and the first to estimate this relationship among primary school students in Africa. Building scientific theories solely on evidence from affluent Western countries will bias empirical evidence, as Western samples are not representative of the general human population. Therefore, the evidence presented here is relevant to the broader international body of research on the role of grit in predicting academic achievement.

Given the major role of poor school quality in depressing learning outcomes in LMICs, an important consideration for international education research and practice is whether the strong association between grit and student achievement observed in high-income countries holds in poor-quality schools. In addition, given the well-documented inequality in school quality that characterises the education systems of LMICs, it is important to investigate whether school characteristics interact with grit to produce learning outcomes. As such, the present study tests for the possibility that school functionality interacts with the perseverance subscale of grit to produce reading outcomes.

3.6.1. SUMMARY OF MAIN FINDINGS

This chapter sought to investigate two main research questions, namely: (1) What relationship, if any, exists between grit and student achievement among Grade 6 learners in a sample of 60 township and rural schools in South Africa? and (2) Is there evidence of moderating effects between school quality and perseverance in predicting learning outcomes? In the analysis of potential interaction effects, the study sought to investigate whether the data supported one of three hypotheses: 1) The Matthew effect hypothesis, whereby perseverance would have a stronger association with reading scores in more functional schools; 2) The independent effects hypothesis, whereby there is no significant difference between perseverance and reading scores across schools with different levels of functionality; and 3) The resource substitution hypothesis, whereby perseverance has a stronger association with reading scores in less functional schools.

In terms of the first research question, the perseverance subscale of grit emerges as a strong predictor of reading achievement among this sample of township and rural schools in South Africa, even when controlling for a host of factors at the individual, home, and school level. Importantly, the perseverance subscale of grit maintains its significance as a predictor of reading achievement when grouping the sample according to levels of school functionality and estimating the association between grit and reading scores separately for each school tercile. That is, I find no evidence to suggest that school functionality limits the power of the perseverance subscale of grit to predict reading outcomes among this sample of schools. These results make an important contribution to the grit literature in that it introduces evidence from high-poverty schools in an African country to the international evidence base of the association between grit and student achievement.

In terms of the second research question of potential interaction effects between grit and school functionality in predicting student achievement, the evidence presented in this chapter leads to two important conclusions. Firstly, the results do provide evidence of such interaction effects, suggesting that the perseverance subscale of grit, specifically, interacts meaningfully with school functionality to predict reading test scores. Secondly, further interrogation of these interaction effects produces evidence that the nature of the interaction between perseverance and school functionality is not uniform across the distributions of these variables. Specifically, the results provide evidence of Matthew effects, resource substitution effects, and independent effects between perseverance and school functionality, depending on where on the distributions of both of these variables one tests for these effects. This result makes an important contribution to the literature on interaction effects between non-cognitive skills and other education inputs, since it provides empirical evidence for the theoretical possibility that the nature of interaction effects between non-cognitive skills and other educational inputs may vary at different points of the distributions of both non-cognitive skills and the inputs they interact with.

3.6.2. LIMITATIONS

Measurement concerns

These results are subject to a number of important limitations. The first major limitation is the low reliability with which grit is measured among the Leadership for Literacy sample. While measures of grit with similarly low levels of reliability have been employed elsewhere in the literature (Light and Nencka, 2019; Wills and Hofmeyr, 2019), the low alpha coefficient of the grit scale is a major limitation of the chapter, since it means firstly that only the perseverance subscale of grit could be used in the analysis, and secondly, even then, we cannot be sure that this measure represents the single latent trait of perseverance. Part of the reason for the low alpha coefficient of the perseverance subscale of grit is likely that it is comprised of only four questionnaire items. An important recommendation for future research is therefore that student surveys should include more items aimed at capturing non-cognitive skills such as perseverance. In addition, findings from the psycholinguistics literature suggest the low internal reliability of the 'consistency of interest' subscale of grit may be due to the fact that the items comprising the scale are negatively worded. Negatively worded items are typically included in questionnaires to guard against acquiescence bias, where respondents tend to agree with statements without regard for their actual content (Salazar, 2015). However, the results presented in this study provide support to growing consensus in the psycholinguistics literature that attempting to correct for acquiescence bias by including negatively worded items introduces artefacts that may be more problematic for construct validity than acquiescence bias (Weems *et al.*, 2003; Weems, Onwuegbuzie and Lustig, 2009; Roszkowski and Soven, 2010; Salazar, 2015). Therefore, the results from this study provide additional support for the recommendation from this literature to not include negatively worded items in educational assessments, especially when student background questionnaires are administered in contexts of low literacy. At the very least, the results presented show that the wording of questionnaire items is extremely important in determining the reliability with which non-cognitive skills are measured and, therefore, requires careful consideration.

The outcome variable of interest used in the analysis, namely scores on a silent comprehension test administered in English, also has limitations as a measure of reading ability, especially since English was a second language for almost all students in the sample (although all of the students have been taught in English since at least Grade 4). Therefore, in addition to focusing future research efforts on measuring non-cognitive skills such as grit more reliably, improving our understanding of the relationship between these skills and student achievement will require devoting significant research efforts towards measuring student achievement more reliably. Incorporating existing evidence from the growing body of evidence from other contexts where students are taught in a second language (such as Mexican children who are taught in English in the US) is likely to be helpful in this regard.

Sampling

The samples of schools in the different functionality terciles are relatively small and were not sampled to be either provincially or nationally representative. Therefore, while the results presented in this study make an important contribution to our understanding of the association between socio-emotional skills and learning outcomes in the South African context, more evidence from different samples of students is required to deepen our understanding of this relationship.

Conceptual limitations

Even if these measurement and sampling issues were addressed, the analysis presented in this chapter is subject to further, more conceptual limitations regarding what is actually being measured in the perseverance subscale of the Grit-S scale. The distinctions between grit and intelligence, as well as that between grit and other non-cognitive skills, are of particular importance in this regard. Regarding the former, since the analysis does not control for intelligence, it might be that the association between the perseverance subscale of grit and reading outcomes found in this study is driven – either partly or entirely – by intelligence (see the work of Light and Nencka (2019); He *et al.* (2021) for evidence regarding the association between grit and intelligence).

Given that the Leadership for Literacy data does not contain measures of intelligence, it is not clear to what extent – if any – differences in intelligence may be affecting the results presented in this chapter. The potential interdependence between grit and intelligence is therefore an important avenue for future research aimed at improving our understanding of the association between grit and student achievement. The first step in this regard would be administering both tests of cognitive ability (such as IQ tests) and grit to the same samples of students. While this would require significant resources and expertise, studies such as the recent work of He *et al.* (2021), which assessed both grit and IQ among just under 3,000 Grade 7 students in rural China, illustrate that much can be gained in terms of our understanding of the relationship between non-cognitive skills such as grit and student achievement in disadvantaged contexts by including measures of cognitive skills in studies that investigate this relationship.

Regarding the distinction between grit and other non-cognitive skills, a number of authors have raised concerns about the discriminant validity of grit – that is, that grit may simply be a new name for psychological constructs that have long been studied, and for which large evidence bases exist, such as self-control, growth mindset, tenacity, determination, and especially the Big Five personality trait of conscientiousness (Dweck, Walton and Cohen, 2011; Ivcevic and Brackett, 2014; Credé, Tynan and

Harms, 2017; Muenks *et al.*, 2017; Credé, 2018; Kannangara *et al.*, 2018). Moreover, studies that explicitly test for the incremental validity of the grit construct show that grit does not add anything to predictions of student achievement when controlling for conscientiousness (Ivcevic and Brackett, 2014; Rimfeld *et al.*, 2016). Based on these results, Rimfeld *et al.* (2016: 786) conclude that "the association between achievement and personality is largely explained by the Big Five and grit adds little to this relationship." In light of these findings, it is likely that what is being measured as "grit" in the Leadership for Literacy data is in fact any of the other non-cognitive skills that overlap with grit. The results presented in this chapter are therefore subject to the important caveat that the choice to investigate the relationship between student achievement and grit specifically was informed largely by data availability. As is the case for disentangling the effects of cognitive skills from non-cognitive skills, disentangling which specific non-cognitive skills are important for predicting student achievement will require devoting significant research effort toward the measurement of more non-cognitive skills in South Africa.

Causality

Lastly, the results presented in this chapter are subject to the usual limitation that causality cannot be inferred from the associations, given that the data is cross-sectional. Studies that attempt to overcome this limitation usually employ early measures of non-cognitive skills to predict later outcomes (Almlund *et al.*, 2011). Indeed, using student-level panel data which includes measures of non-cognitive skills to estimate causal effects of non-cognitive skills on student achievement is a fast-growing area of research within the economics of education (see for example Cunha and Heckman (2007, 2008); Agostinelli (2018); Agostinelli, Saharkhiz and Wiswall (2019); Attanasio *et al.* (2020)). It is important to note, however, that the application of these techniques to contexts such as South Africa will not be possible as long as such panel data is not available. As a result, it is imperative that in addition to directing efforts towards the measurement of IQ and multiple measures of non-cognitive skills, efforts are made to collect panels of these measures. Only when such data becomes available will it be possible to attempt to infer causality in the relationship between non-cognitive skills and academic achievement for South African children.

3.7. CONCLUSION

This study contributes to the international evidence base of the association between grit and student achievement by estimating this relationship for a sample of Grade 6 students in township and rural schools in South Africa. The analysis in this chapter was aimed at investigating the potential interdependence between grit and school functionality in predicting learning outcomes. The results suggest that the perseverance subscale of grit is a strong predictor of reading achievement, even for student who attend schools with very low levels of functionality. In addition, the results are suggestive not only of significant interaction effects between perseverance and school functionality in predicting reading achievement, but also of variation in the nature of these interaction effects for students at different points of both the perseverance and school functionality distributions. These results make an important contribution to the international evidence base of the association between grit and student achievement, and add to our understanding of how non-cognitive skills such as grit may interact with school characteristics in predicting learning outcomes. As such, investigating how other non-cognitive skills interact with school characteristics in producing learning outcomes is likely to be a fruitful avenue for future research.

CHAPTER 4: SOUTH AFRICA'S PRO-GIRL GAP IN PIRLS AND TIMSS: HOW MUCH CAN BE EXPLAINED?

4.1. INTRODUCTION

Increasing girls' participation in education has been a central feature of efforts to promote educational equality in LMICs. What has received less attention in both the literature and policy debates is that gender issues in education vary considerably across the developing world. One major area of divergence is that while some developing countries are still struggling to achieve gender parity in school enrolment (Grant and Behrman, 2010), others exhibit the same pro-girl advantage in education as is observed in most of the industrialised world (Badr, Morrissey and Appleton, 2012). As shown in the previous two chapters, South Africa is one such a LMIC where girls systematically achieve better educational outcomes than boys. Girls achieve better results in virtually all the international educational assessments South Africa participates in (Van Broekhuizen and Spaull, 2017; Zuze *et al.*, 2017; Spaull and Makaluza, 2019), are less likely to repeat a grade and drop out of school (Van der Berg *et al.*, 2019), and achieve better results in *matric* – the school-leaving exam in South Africa (Spaull and Makaluza, 2019). This pro-girl achievement gap in basic education continues through higher education, where women are 34% more likely to enrol in undergraduate programmes than men on average, and 62% more likely to complete their undergraduate degrees than men (Van Broekhuizen and Spaull, 2017). While a pro-female advantage in education is not in itself unusual - most high-income countries exhibit similar pro-female gender gaps in basic and higher education (Jacob, 2002) - South Africa's pro-girl achievement gap is noteworthy for at least two reasons: (i) The fact that such a distinct and persistent gap exists in an education system that is in many ways more similar to those in LMICs; and (ii) The magnitude of the gap, especially in the primary school grades - the pro-girl advantage in Grade 4 reading achievement is roughly *four times* as large as that observed in other countries[21].

Although these features of South Africa's pro-girl achievement gap are in themselves noteworthy, these country-level averages mask important differences in the magnitude and extent of the gender achievement gap across socio-economic and schooling contexts in South Africa. The handful of local studies that investigate differences in the magnitude of the gender achievement gap in South Africa present evidence to suggest that gender and SES intersect in meaningful ways to produce learning outcomes (Van Broekhuizen and Spaull, 2017; Zuze and Beku, 2019), a result which has largely been overlooked in local research, at least compared with the international literature, where the intersection between SES and gender in influencing educational outcomes has received considerable attention.

[21] This is according to South Africa's PIRLS (2016) results. South Africa's pro-girl advantage is roughly four times as large as the average pro-girl gap across all participating countries (Mullis *et al.*, 2017).

Another feature of the pro-girl achievement gap in South Africa that has received little attention is potential sources of this gap. While the international literature has been concerned with identifying the sources of the pro-girl advantage in educational outcomes for more than two decades, local studies have generally focussed more on documenting gender gaps in education rather than attempting to provide explanations for them (see Van Broekhuizen and Spaull (2017); Spaull and Makaluza (2019); Zuze and Beku (2019)). The aim of this study is therefore to interrogate potential reasons for the pro-girl achievement gap in South Africa. Specifically, given mounting evidence from the international literature of the importance of gender differences in non-cognitive skills – such as attitudes toward school and learning - in explaining gender achievement gaps, the analysis in this study is aimed at examining the extent to which potential differences between boys and girls in these skills can explain the pro-girl achievement gap in South Africa.

The main econometric strategy employed to this end is Oaxaca-Blinder decomposition analysis, whereby the pro-girl gap in achievement is split into two components - that which can be explained due to gender differences in observable characteristics, and that which remains unexplained. The decomposition analysis is performed separately on South African data from two international educational assessments, namely the Progress in International Reading Literacy Study (PIRLS), which tests reading in Grade 4, and the Trends in Mathematics and Science Study (TIMSS), which tests mathematics in Grade 5.[22] Importantly, given evidence of gender differences in grade repetition in Grades 1-3, I perform two separate sets of decomposition analyses - one on the full PIRLS and TIMSS samples, and one on restricted samples of these datasets where I attempt to control for gendered repetition patterns in prior grades. This is done in order to determine whether other observable differences between boys and girls – apart from being overage – contribute to the observed pro-girl achievement gaps. Given evidence from the international literature that the magnitude and potential sources of gender achievement gaps tend to differ by SES, each decomposition analysis is performed separately by school quintile within each dataset. This approach allows me to investigate potential SES differences in the magnitude and factors contributing to South Africa's pro-girl gap in both Grade 4 reading and Grade 5 mathematics achievement.

This chapter makes four important contributions to the literature on South Africa's pro-girl advantage in educational outcomes. Firstly, I show that around half of the pro-girl achievement gaps in PIRLS and TIMSS can be explained by observable differences between boys and girls. While this explained proportion is comparable to studies conducted in other countries that use a similar methodology, this result still means that around half of the pro-girl achievement gap remains unexplained.

[22] While most countries participate in the TIMSS study in Grade 4, countries can choose to participate in Grade 5 if they suspect that the assessment will be too difficult for Grade 4 students.

Secondly, the results from the decomposition analysis suggest part of the observed pro-girl achievement gap in both PIRLS and TIMSS is explained by boys being much more likely to be overage for their grade than girls. Since being overage is an indicator of having repeated a grade, this implies that any given Grade 4 or 5 class will consist of a larger proportion of boys than girls who have repeated an earlier grade. In this sense, boys in these grades are already “selected” to be weaker performers than their female peers. While one may expect grade repetition to be associated with learning gains, whereby repeaters “catch up” their initial backlog in learning, existing evidence suggests that grade repetition does not achieve this aim in South Africa, since repeating a grade is associated with weaker subsequent performance (Anderson, Case and Lam, 2001; Van der Berg *et al.*, 2019). Part of the observed pro-girl achievement gaps in PIRLS and TIMSS can therefore be attributed to a female advantage that is already evident at the start of formal schooling and accumulates over the early grades. This result makes an important contribution to the literature on gender gaps in achievement in countries with high rates of grade repetition such as South Africa.

Thirdly, I present evidence of gender gaps in non-cognitive skills, specifically attitudes towards school and the learning process, and how these differences contribute to the pro-girl achievement gap in South Africa. Notably, I find that gender differences in non-cognitive skills may be an important contributing factor to South Africa’s pro-girl advantage in PIRLS. This result is in accordance with findings from the international literature, and highlights the need for more research on the role of non-cognitive skills in contributing to South Africa’s pro-girl advantage in reading achievement. Lastly, performing the decomposition analysis separately by school quintile allows me to uncover SES differences in both the magnitude and factors contributing to the pro-girl achievement gaps in PIRLS and TIMSS. I find that after accounting for students’ age, the pro-girl achievement gap decreases with school socio-economic status, and the factors that contribute to the pro-girl gap differ across school socio-economic status.

The chapter is structured as follows: Section 4.2 provides a brief review of the international literature on pro-girl achievement gaps, paying particular attention to studies that investigate the role of gender differences in student attitudes toward learning and school. Details of the estimation samples and key measures used in the decomposition analysis are presented in Section 4.3. Section 4.4 documents the magnitude of the pro-girl achievement gap across school quintiles in both PIRLS and TIMSS, and provides descriptive evidence of how this gap is related to gendered grade repetition patterns in the foundation phase (Grades 1-3). The results of the decomposition analysis which investigates potential sources of the pro-girl gap among students who are on-track in terms of age-for-grade are discussed in Sections 4.5 and the implications of these results are discussed in Section 4.6. Section 4.8 concludes.

4.2. LITERATURE REVIEW

There is a vast literature demonstrating that learning outcomes are surprisingly gendered. Evidence from industrialised countries points to remarkable uniformity in gender achievement gaps in the primary school grades, which can be summarised quite simply: girls do better in reading and boys do better in mathematics (Cobb-Clark and Moschion, 2017). The evidence from LMICs, on the other hand, is more difficult to summarise, given much more variation in the magnitude and direction of gender achievement gaps across countries. For example, in a comparison of gender achievement gaps in four LMICs, Dercon and Singh (2013) find that Ethiopia and India exhibit a pro-boy advantage in Grade 6 mathematics achievement, while Vietnam exhibits a distinctly pro-girl advantage, and Peru shows no significant gender gap in mathematics achievement at this age. Similarly, Zuze (2015) presents evidence of variation in the magnitude of gender achievement gaps within the East African region, with Kenya and Tanzania exhibiting large pro-boy gaps in SACMEQ[23] Grade 6 mathematics achievement, and Uganda not exhibiting significant gender differences in achievement in the same assessment. Based on this evidence, Zuze (2015) concludes that gender issues are unique to countries, even those in the same geographic region, and the factors that contribute to gender achievement gaps are largely determined by the local context, both between and within countries.

Part of the difficulty in interpreting evidence of achievement gaps in developing countries lies in correcting for sample selection bias that results from gender differences in access to schooling. While major progress has been made in terms of ensuring equal access to schooling – especially since reducing gender gaps disadvantaging girls became part of the UN's Millennium Development Goals in 2000 (Grant & Behrman, 2010) − in many developing countries access is still biased against girls. When analysing gender gaps in achievement, one therefore has to account for these differential access rates (see for example Grant & Behrman, 2010). While this is not the case in South Africa, which exhibits near-universal enrolment during the compulsory phase of schooling (that is, up to Grade 9) (Van der Berg and Hofmeyr, 2018), there is a different source of sample selection bias that may exaggerate observed gender gaps in achievement, namely differential rates of grade repetition by gender. Van der Berg *et al.* (2019) present a comprehensive account of grade repetition in South Africa, and find that boys are more likely than girls to repeat in virtually all grades. Importantly, they present evidence that boys' disadvantage in terms of grade completion already begins in Grade 1, and only becomes more pronounced as students progress through school. This results in more boys being overage – that is, older than they should be in a particular grade – than girls in almost all grades. Given that grade repetition is usually associated with weaker academic performance (Ikeda and García, 2014; Sunny *et al.*, 2017), gendered patterns of grade repetition in Grades 1-3 may cause a selection effect whereby a larger proportion of boys than girls in a given Grade 4 or 5 class are already "selected" to be weaker-performing, which would exaggerate pro-girl achievement gaps in PIRLS and TIMSS. In essence, this would mean that girls already outperform boys in the foundation phase and that the achievement gaps

[23] The Southern and Eastern African Consortium for Monitoring Educational Quality, a standardized assessment conducted in 15 countries in southern and eastern Africa.

observed in PIRLS and TIMSS are largely the result of a pro-girl advantage that accumulates over the course of the early grades. While investigating the sources of this gap in the early grades is beyond the scope of this chapter, it is still useful to know how much of the observed pro-girl achievement gaps in PIRLS and TIMSS can be attributed to these gendered repetition patterns in the early grades. This question is therefore explored explicitly in the decomposition analysis.

The literature distinguishes between four broad categories of explanations for gender gaps in educational achievement (Wilsenach and Makaure, 2018): (i) biological, where girls' superior academic achievement is linked to evidence that girls develop the cognitive skills that underpin learning earlier than boys (Gierl *et al.*, 2003; Rosselli *et al.*, 2009; Andreoni *et al.*, 2019); (ii) parents' gender-specific expectations and investments (Kingdon, 2002; Entwisle, Alexander and Olson, 2007; Mencarini, Pasqua and Romiti, 2019); (iii) schooling, where it is argued that educational practices favour girls (Entwisle, Alexander and Olson, 2007); and (iv) gender differences in the acquisition of non-cognitive skills that support learning. Mounting evidence of the important role that non-cognitive skills play in determining educational outcomes more generally (Heckman, 2000; Almlund *et al.*, 2011; Heckman and Kautz, 2012; Diaz, Arias and Tudela, 2013; Garcia, 2013; Kautz *et al.*, 2014; Stankov and Lee, 2014; Egalite, Mills and Greene, 2016) has led to increased attention to potential gender differences in these skills in the achievement gap literature.

Existing evidence points to a pro-girl gap in student attitudes, including subject-specific self-efficacy beliefs (Kennedy, 2008; Popp *et al.*, 2014; McGeown *et al.*, 2015), engagement in lessons (Van de Gaer *et al.*, 2009; DiPrete and Jennings, 2012; McGeown *et al.*, 2015), reading enjoyment (Kennedy, 2008; Logan and Johnston, 2009; Marinak and Gambrell, 2010; Mol and Jolles, 2014), and mathematics enjoyment (Hemmings, Grootenboer and Kay, 2011). Studies in this strand have also considered gender differences in students' sense of belonging at school and the frequency with which bullying is experienced at school as potential sources of the pro-girl achievement gap, with existing evidence pointing to girls having a higher sense of school belonging (Goodenow, 1993; Sánchez, Colón and Esparza, 2005; Hughes, Myung and Allee, 2015) and being less likely to experience bullying at school than boys (Scheithauer *et al.*, 2006; Popp *et al.*, 2014).

It is important to note, however, that these distinctions between categories of explanations for gender gaps in educational outcomes - usually made for methodological reasons - are largely artificial in the sense that learning outcomes result from a combination of "overlapping spheres of influence" (Alexander, 2016: 18), making it nearly impossible to disentangle the multitude of factors that contribute to the gender achievement gap (DiPrete and Jennings, 2012; Cobb-Clark and Moschion, 2017). Importantly, two decades of research on the formation of skills (Heckman, 2006) has shown that developmental environments are crucial in shaping both cognitive and non-cognitive skills. This finding blurs the distinctions between the four categories of explanations for the gender achievement gap listed above, since we now know that children's cognitive (explanation (i)) and non-cognitive skills (explanation (iv)) are shaped jointly by genetics, their environments (explanation (ii)), and parental investments (explanation (iii)).

A number of authors have attempted to investigate how different spheres of influence interact to produce gender gaps in educational outcomes by evaluating these gaps separately for students at different points of the SES distribution. The consensus based on evidence from high-income countries seems to be that pro-female gaps in educational outcomes are more pronounced among low-SES students (Bertrand and Pan, 2011; Legewie and DiPrete, 2012; Autor *et al.*, 2016; Mencarini, Pasqua and Romiti, 2019), with some studies finding evidence to suggest that pro-girl advantages in achievement are found *only* among socioeconomically disadvantaged students (see Entwisle, Alexander and Olson (2007) for a review of this literature). A number of potential reasons have been proposed for this evidence that gender achievement gaps tend to be more pronounced among socioeconomically disadvantaged children. These can broadly be grouped into two categories, namely (i) explanations that ascribe this phenomenon to SES differences in parents' and teachers' expectations of boys and girls, with parents and teachers tending to have lower expectations of low-SES boys – an effect which is absent among high-SES students (Entwisle, Alexander and Olson, 2007; Strand, 2010), and (ii) explanations that hinge on the notion that characteristics of boys' and girls' home and school environments are translated into educational outcomes in different ways. In particular, a number of authors present evidence to suggest that boys' educational outcomes are more sensitive to socioeconomic home disadvantage (Bertrand and Pan, 2011; Marcenaro–Gutierrez, Lopez–Agudo and Ropero-García, 2018; Mencarini, Pasqua and Romiti, 2019) and low-quality schools (Legewie and DiPrete, 2012; Autor *et al.*, 2016).

In summary, there are three important findings from the international literature regarding the nature and potential sources of gender gaps in educational outcomes that inform the research design of this study. Firstly, gender gaps differ across subjects, with girls typically outperforming boys in reading, while boys tend to have an advantage in mathematics. Secondly, there is evidence to suggest that gender gaps differ by SES, something that has received little attention in the South African literature on gender gaps in educational outcomes. Thirdly, potential gender differences in student attitudes are a likely source of South Africa's pro-girl gender gap in educational outcomes that have not yet been investigated in the local literature.

The aim of this study is thus to contribute to our understanding of the magnitude and potential sources of the pro-girl achievement gap in South Africa by examining variation in this gap across school quintiles in two subjects, namely reading and mathematics, and to investigate whether student attitudes contribute to this gap. Evidence from the local literature suggests, however, that there may be gendered sample selection processes underlying the composition of any given class participating in international assessments such as PIRLS and TIMSS. Specifically, the fact that boys are more likely to be held back in the foundation phase may mean that the samples of boys participating in these assessments are already "selected" to be weaker-performing than girls. The analysis in this chapter therefore attempts to answer two questions, namely (i) How much of the observed pro-girl achievement gaps in PIRLS and TIMSS can be attributed to gendered repetition patterns in the early grades?; and (ii) How much of this achievement gap can be attributed to observable differences between boys and girls, when comparing only boys and girls who are on-track in terms of age-for-grade?

4.3. DATA

I make use of the most recent publicly available data from two international educational assessments that South Africa participates in, namely the 2016 round of the Progress in International Reading Literacy Study (PIRLS) for developing countries, PIRLS Literacy, and the Trends in Mathematics and Science Study (TIMSS) study for developing countries, TIMSS Numeracy, which was conducted in 2015.

4.3.1. ESTIMATION SAMPLES

PIRLS Literacy 2016

PIRLS is an international large-scale literacy assessment conducted by the International Association of the Evaluation of Educational Achievement (IEA). PIRLS Literacy 2016 was administered by the Centre for Evaluation and Assessment (CEA) at the University of Pretoria. PIRLS 2016 employed a two-stage stratified cluster sampling design so that a nationally representative sample of schools was chosen according to province and the school's language of instruction in the foundation phase (Howie *et al.*, 2017). Within the sampled schools, classes were randomly selected for participation. Sampled classes thus constitute the second-stage sampling units. All students in sampled classes present on the day of the assessment participated in the assessments. In 2016 the realised PIRLS sample consisted of 12,810 Grade 4 students from 293 schools across South Africa. An interesting feature of South Africa's PIRLS data that has not received attention is the fact that the sample is not perfectly balanced in terms of gender – boys make up 51.8% of the sample. Unfortunately, the data contains some missing information in the student attitude measures. Since these variables constitute some of the main covariates of interest, students who had missing information on these measures were dropped from the sample. This resulted in 1,027 students (8% of students) being dropped from the original PIRLS sample. The sample used in my analysis therefore consists of 11,734 students from all 293 schools. Boys were slightly more likely to have missing information on the student attitude measures, resulting in boys comprising 51.2% of the final estimation sample.

TIMSS Numeracy 2015

TIMSS Numeracy is also conducted by the IEA and was administered in South Africa by the Human Sciences Research Council (HSRC). TIMSS collects the same contextual information from students, parents, teachers and principal as PIRLS (Human Sciences Research Council, 2017). Students were sampled using the same two-stage stratified cluster sampling design as employed in PIRLS. The TIMSS 2015 realised sample consists of 10,932 Grade 5 students from 297 schools (51.6% male). The same missing data concerns present in PIRLS plague the TIMSS data, thus students with missing information on the student attitude variables were also dropped from the original TIMSS sample. This resulted in the deletion of 649 observations (6% of the original sample). The TIMSS sample used in the analysis that follows thus consists of 10,283 Grade 5 students (51.6% male) from all 297 schools.

4.3.2. DESCRIPTION OF MEASURES

(I) EDUCATIONAL ACHIEVEMENT MEASURES

The PIRLS Literacy assessment consisted of a silent reading comprehension test administered in all South Africa's 11 official languages. The school language policy of South Africa is currently implemented in such a way that the language of learning and teaching (LOLT) for the vast majority of students is their home language in Grades 1-3, and from Grade 4 there is a LOLT switch to English for the remaining school years (Spaull and Kotze, 2015). By Grade 4, the majority of South African students would have had limited exposure to English, and consequently the comprehension test was administered in the school's LOLT in the foundation phase. The TIMSS Numeracy assessment tested fundamental mathematical knowledge, procedures, and problem-solving skills (Human Sciences Research Council, 2017) and was administered in English or Afrikaans. Both PIRLS and TIMSS test scores are standardised to have an international mean of 500 points and a standard deviation of 100 points, and these scores were standardised for each sample used in my analysis to have a mean of zero and a standard deviation of one for ease of interpretation of the multivariate results.

It is important to note that the PIRLS and TIMSS assessments differ markedly in terms of their relative difficulty. PIRLS was developed for a predominantly high-income country context (Spaull and Pretorius, 2019) and it is clear from South Africa's results that students found this assessment very challenging, with only 22% of students reaching the low international benchmark of 400 points. TIMSS Numeracy, on the other hand, was also developed for predominantly high-income countries, but countries could opt to administer the assessment to Grade 5 students, if they suspected that it would be too difficult for Grade 4 students. It is clear from South Africa's overall performance in TIMSS that students found the assessment less challenging than PIRLS, with 40% of students reaching the low international benchmark, though South African students were tested in Grade 5.

(II) SCHOOL SES MEASURE

Following Spaull and Pretorius (2019), I calculate school SES as the average SES of all students in a school. Student SES was measured using information about eight possessions students indicated having in their homes. Principal components analysis (PCA) was used to derive an index from these variables in the PIRLS and TIMSS samples, respectively[24]. As Spaull and Pretorius (2019) point out, this method is unlikely to create an accurate cardinal measure of wealth. However, since my purpose in creating an asset index is the same as theirs – to create an ordinal ranking of student wealth – I maintain, as they do, that an asset index created from the home possessions variables in PIRLS and TIMSS is the best measure of student wealth available. The decision to use an asset index to measure school wealth in PIRLS, and not test language as in Chapter 2, was informed by the aims of this chapter, namely to examine variation in the pro-girl gap by school SES. This objective means that the measure of school wealth can be used to split the sample into smaller groups than the measure used in Chapter 2, where the objective is to identify academically resilient students from poor schools. In that chapter, using the language the PIRLS test was written in suffices as a proxy for school wealth, since the analysis of academic resilience does not involve examining variation within the poorer part of the school system. The PIRLS asset index has an alpha coefficient of 0.61, and the TIMSS asset index has an alpha coefficient of 0.63. Average school SES was calculated as the mean of this asset index at the school level, and this variable was used to split the schools in each sample into SES quintiles, from the poorest 20% of schools (Quintile 1) to the wealthiest 20% of schools (Quintile 5).

(III) STUDENT ATTITUDE MEASURES

In addition to student assessment data, PIRLS and TIMSS also administered student background questionnaires that included a section aimed to measure student engagement and attitudes toward reading and mathematics, respectively. Students had to choose from a four-point Likert scale (from "disagree a lot" to "agree a lot") on a number of items aimed at quantifying three specific constructs, namely confidence in reading/mathematics, engagement in reading/mathematics lessons, and reading/mathematics enjoyment. Indices of each of these constructs were then created by aggregating across the different items that were intended to capture each construct, with scores on each index ranging between 1 (low) and 3 (high). Student questionnaires also included items aimed at quantifying students' perceptions of school climate and safety, with specific emphasis on students' sense of belonging at school and student bullying. Items included to measure student's engagement and attitudes and their sense of belonging at school were also measured with four-point Likert items (from "disagree a lot" to "agree a lot"). Scores on these items were aggregated to create an index of school belonging ranging from 1 (low) to 3 (high)[25]. Student bullying was measured with eight items that asked students to report how often they had experienced different types of bullying, with response options "Never or almost never", "A few times a year", "Once or twice a month", and "At least once a week"[26]. Scores on

[24] A scree plot of the Eigenvalues of the PCA's can be found in Appendix C (Figures C1 and C2).

[25] See http://timssandpirls.bc.edu/pirls2016/international-results/pirls/student-engagement-and-attitudes/ for details about the procedures followed when assigning scores to the student attitude variables.

[26] Responses to the individual questionnaire items, by gender and students' age, can be found in Tables A3-A13 of the Appendix.

these items were also added up and converted to an index with values ranging from 1 to 3. Thus a score of 1 translates to "never or almost never", 2 infers "about monthly", and 3 infers "about once a week". For ease of interpretation of the multivariate results, I standardised the student attitude, engagement, belonging, and bullying indices to have a mean of zero and a standard deviation of one.

(IV) GRADE REPETITION MEASURE

Unfortunately, neither the PIRLS and TIMSS data include direct measures of whether a student has repeated a grade. As such, I use students' age as a proxy for grade repetition. I code this as a dummy variable which takes a value of 1 if a student is one or more years older than they would be if they were "on-track" in terms of age-for-grade.

(V) SCHOOL CHARACTERISTICS

The decomposition models include only two variables that are intended to capture school characteristics that should theoretically matter for student achievement, namely dummy variables indicating whether the school has a library and whether the school has at least one computer. This decision was informed by two considerations. Firstly, there is not enough overlap between the school and classroom characteristics captured in South Africa's PIRLS 2016 and TIMSS 2015 data, respectively, to allow for meaningful comparison of the relative contributions of the various school characteristics in each dataset to the observed gender gaps in achievement. Since one of the aims of this study is to compare the relative importance of specific observable differences between boys and girls in explaining the pro-girl achievement gaps in these datasets, a decision was made to only include those school-level factors that are available in both datasets (namely the library and computer dummies). Secondly, I maintain that, given the high degree of overlap between student SES and school resources in South Africa (Spaull, 2013), meaningful information about school context is captured by my variable of school socio-economic status. That is, while I do not explicitly control for school characteristics in the multivariate analysis (apart from the library and computer dummies), variation in school characteristics is implicitly controlled for through the splitting of the samples into school quintiles and performing the decomposition analysis separately by school quintile.

(VI) OTHER COVARIATES

In addition to student attitudes and age, I include a number of measures to capture individual characteristics and behaviours of students, including whether they attended preschool, the frequency with which they did homework (parent-reported), student SES (measured using the asset index described above), and province. Given a large extent of missing information on the parent-reported homework frequency variable, this variable was recoded as a dummy indicating whether parents reported their child did homework at least three times a week or not. Missing values were grouped with the reference category. Lastly, I include a dummy variable indicating whether the student wrote the test in their first language. This variable was derived from the student background questionnaire item that asked students about the frequency with which they spoke the language of the test at home. Students who answered “always” or “almost always” on this question were assigned a value of 1 on the “first language” variable.

4.3.3. DESCRIPTIVE DIFFERENCES BETWEEN BOYS AND GIRLS

Table 14 and Table 15 show differences between boys and girls in these observable characteristics, by school quintile, for the PIRLS and TIMSS samples, respectively. Table 14 presents evidence of statistically significant differences between boys and girls in PIRLS in all the observable characteristics included here, across virtually all school quintiles. The results that girls are more confident in their reading abilities, report more engaging teaching by their reading teachers, enjoy reading more, report a higher sense of school belonging, and report experiencing bullying less often than boys are in accordance with the evidence presented in a number of existing studies, as discussed in Section 4.2. The fact that girls do homework more often than boys (according to parent reports) suggests that in addition to girls having more positive attitudes toward reading and school more generally, girls are also more likely to exhibit behaviour that supports learning – such as doing homework.

In accordance with the existing literature, the results in Table 14 indicate that boys are more likely to be overage than girls in all school quintiles. In addition, the results in the table point to large differences in the proportions of overage students across school quintiles, with larger proportions of overage students in poorer schools. This result is in line with the findings of Van der Berg *et al.* (2019) that grade repetition is more common in the lower school quintiles. Importantly, the proportions of overage boys and girls in each sample roughly match repetition rates in the early grades reported in existing studies (see Branson, Hofmeyr and Lam, 2014). This constitutes evidence that being overage is a good proxy for grade repetition in the PIRLS and TIMSS samples.

The results in Table 14 also show that girls have significantly lower asset index scores than boys, that is, that girls come from poorer homes than boys, on average. This is an unexpected result since gender is usually assumed to be exogenously determined (Cobb-Clark and Moschion, 2017). The reason for the apparent pro-boy advantage in wealth most likely lies in the fact that some of the assets included in the PIRLS student background questionnaire may have a gendered dimension, whereby parents may be more likely to acquire certain assets, depending on the gender of their child. Table B1 in Appendix B reports gender differences in each of the home assets, and provides evidence that boys are more likely to report having almost all the assets included in the PIRLS student background questionnaire in the homes, with boys being especially more likely to report having a gaming station, internet access, and their own bedroom. Conversely, girls are more likely to report having a study desk and their own story books. These gender differences must be interpreted with caution, however, since the asset measures are likely to be subject to response bias, and, moreover, since the extent of this bias may differ by gender (for example, Engzell (2019) presents evidence of gender differences in the likelihood of over-reporting the number of books in the home). The fact that boys in PIRLS scored higher on the asset index could therefore indicate that the assets included in the student background questionnaire are not gender neutral, or that boys are more prone to overreporting on certain assets, or both[27].

Interestingly, there appears to be less stark gender differences in students' attitudes toward mathematics than is the case for reading: Table 15 shows that girls only consistently scored higher than boys in all school quintiles on the variable measuring students' sense of belonging at school – a general measure of student belongingness, which is not mathematics-specific. The lack of statistically significant gender differences in mathematics self-concept ("Confidence index") in four out of the five school quintiles is particularly noteworthy, since subject-specific self-concept is usually strongly related with achievement. It is noteworthy that girls are no more confident in their mathematics abilities than boys, despite outperforming boys in mathematics in almost all quintiles. The fact that girls in the South African TIMSS are not significantly more confident in their mathematics abilities despite outperforming boys in the assessment accords with evidence in the international literature (Tapia and Marsh, 2000; Bofah and Hannula, 2015), and points to potential internalised gender stereotypes that boys are "naturally" better at mathematics (Steele, 1997).

As is the case in PIRLS, girls in TIMSS report being bullied significantly less often than boys in all school quintiles, and do homework more often than boys, according to parent reports. Girls in TIMSS are also much less likely to be overage than boys, a result that is consistent across school quintiles. The same SES differences in the extent of overage students observed in PIRLS are observable in TIMSS, whereby students in poorer schools are much more likely to be overage than their counterparts in wealthier schools. Once again, this result is consistent with Van der Berg *et al.*'s (2019) finding that grade repetition is more prevalent in the lower quintiles. It is further interesting to note from Table 15 that the same pro-boy advantage in student wealth observed in PIRLS is not observable in TIMSS, where only boys in the bottom school quintile scored significantly higher than girls on the asset index[28]. This result reflects the fact that the assets included in the TIMSS student background questionnaire are less likely to differ by gender, and in this sense the asset index constructed from the TIMSS home possession variables is likely to be a better measure of student wealth than the asset index constructed from the PIRLS home possession variables. However, I maintain that both asset indices suffice as ordinal measures of student wealth[29].

---

[27] Gender differences in student-reported home possessions should not bias the measures of school SES, since there are roughly equal numbers of boys relative to girls in each school quintile. Any gender bias in either the presence of certain assets in students' homes, or the extent of over- or under-reporting on these asset measures, is therefore expected to be uniform across school quintiles. The result that there are significant gender differences in the home possession variables in both PIRLS and TIMSS does, however, constitute a noteworthy finding that should be interrogated in future research.

[28] See Table A2 in the Appendix for responses to individual home possessions, by gender and school quintile.

[29] See footnote 27.

The results in Table 15 further point to a clear SES dimension to gender differences in student attitudes in TIMSS, with girls in poorer schools generally scoring higher than boys on the student attitude measures, while there are fewer significant gender gaps in student attitudes in wealthier schools. This is an unexpected result which points to potential SES differences in the nature of the pro-girl gap in TIMSS mathematics achievement. This result is explored further in the decomposition analysis. Considered together, the results from Table 14 and Table 15 point to an interesting result, namely that while girls have much more positive attitudes towards reading than boys, the same is not generally true for girls' attitudes towards mathematics. This constitutes a noteworthy finding, since it points to an element of domain specificity to gender differences in student attitudes toward learning, whereby the magnitude of these differences depends in part on the subject which is being assessed. This result is also explored further in the decomposition analysis.

Table 14: Descriptive differences between boys and girls, by school quintile – PIRLS (Grade 4)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5		Total	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Confidence index	-0.268	-0.065***	-0.110	0.025***	-0.111	0.064***	-0.034	0.149***	0.102	0.314***	-0.078	0.104***
Engagement index	-0.159	-0.009***	-0.184	0.070***	-0.091	0.114***	-0.034	0.209***	0.036	0.247***	-0.083	0.132***
Enjoyment index	-0.190	0.051***	-0.133	0.097***	-0.103	0.195***	-0.013	0.270***	-0.158	0.187***	-0.116	0.167***
Belonging index	-0.006	0.096**	-0.092	0.128***	-0.142	0.083***	-0.053	0.147***	-0.084	0.122***	-0.076	0.116***
Bullying index	-0.030	-0.126**	0.134	0.000***	0.171	-0.030***	0.071	-0.07***	-0.090	-0.237***	0.055	-0.093***
Overage	0.404	0.261***	0.412	0.240***	0.406	0.263***	0.388	0.248***	0.322	0.206***	0.386	0.243***
Homework	0.277	0.338***	0.379	0.486***	0.342	0.452***	0.250	0.346***	0.400	0.500***	0.331	0.426***
Asset index	-0.758	-0.800	-0.210	-0.341***	0.084	-0.076***	0.392	0.241***	0.856	0.708***	0.103	-0.022***
N	1,037	987	1,171	1,085	1,322	1,238	1,251	1,252	1,228	1,163	6,009	5,725
Proportion of N	51%	49%	52%	48%	52%	48%	50%	50%	51%	49%	51%	49%

Sources: Author's calculations from PIRLS 2016 (Reduced sample of 11,734 students (51% male)). Notes: All indices are standardised to have a mean of zero and a standard deviation of one. Asterisks indicate statistically significant gender differences at *p<0.10, **p<0.05. ***p<0.01. 'Homework' is a dummy variable indicating the proportion of students whose parents reported that they do homework at least three times a week.

Table 15: Descriptive differences between boys and girls, by school quintile – TIMSS (Grade 5)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5		Total	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Confidence index	-0.212	-0.154	-0.211	-0.092***	-0.063	-0.028	0.065	0.048	0.386	0.325	-0.015	0.015
Engagement index	-0.230	-0.086***	-0.298	-0.100***	-0.019	0.057*	0.079	0.165**	0.209	0.239	0.053	0.055***
Enjoyment index	-0.247	-0.143**	-0.208	-0.041***	0.031	0.069	0.078	0.147	0.137	0.149	-0.038	0.040***
Belonging index	-0.193	-0.033***	-0.206	-0.020***	-0.066	0.068***	0.067	0.152**	0.053	0.175**	-0.067	0.069***
Bullying index	0.179	0.054**	0.317	0.111***	0.130	-0.080***	0.000	-0.169***	-0.225	-0.388***	0.088	-0.090***
Overage	0.425	0.267***	0.376	0.190***	0.408	0.253***	0.302	0.186***	0.244	0.119***	0.352	0.204***
Homework	0.520	0.550	0.554	0.630****	0.500	0.581***	0.581	0.675***	0.676	0.749***	0.562	0.636***
Asset index	-0.533	-0.635**	-0.169	-0.209	0.062	0.021	0.312	0.326	0.791	0.821	0.091	0.060
N	904	898	1,089	1,051	1,156	1,129	1,187	1,093	882	894	5,218	5,065
Proportion of N	50%	50%	51%	49%	51%	49%	52%	48%	50%	50%	51%	49%

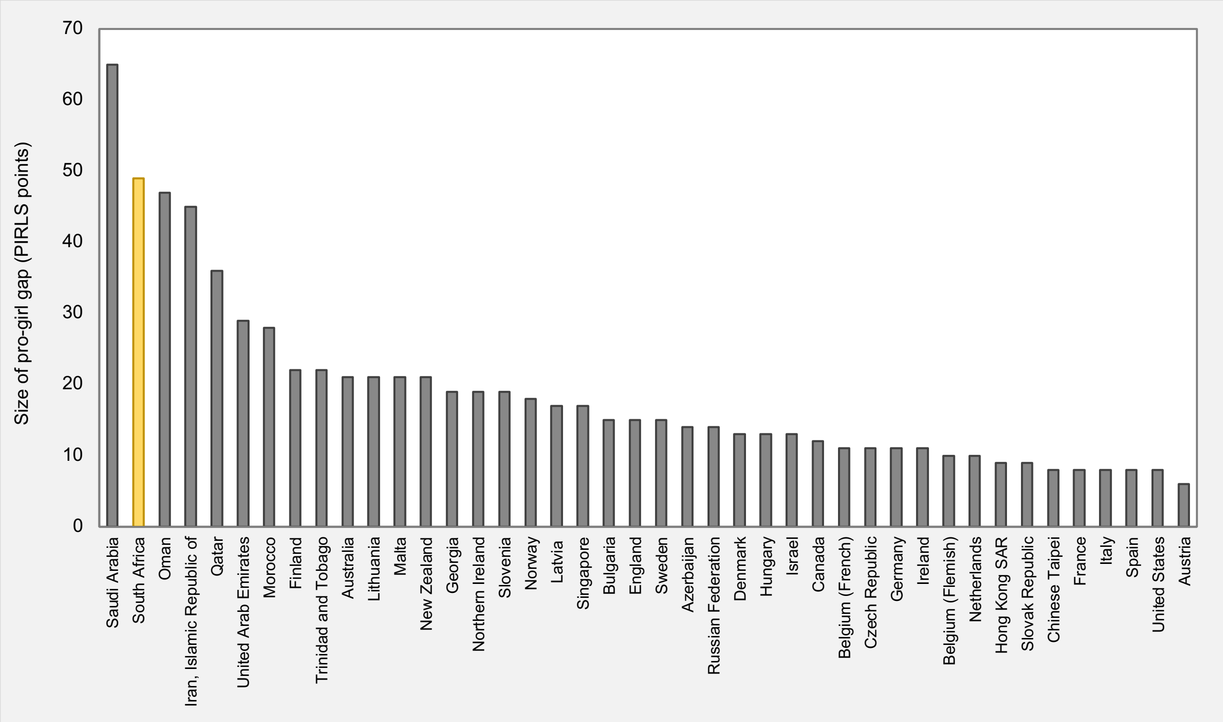
Sources: Author's calculations from TIMSS Numeracy 2015 (Reduced sample of 10,283 students (51% male)). Notes: All indices are standardised to have a mean of zero and a standard deviation of one. Asterisks indicate statistically significant gender differences at *p<0.10, **p<0.05. ***p<0.01. 'Homework' is a dummy variable indicating the proportion of students whose parents reported that they do homework at least three times a week.

4.4. THE MAGNITUDE OF THE PRO-GIRL ACHIEVEMENT GAP IN READING AND MATHEMATICS

The magnitudes of the pro-girl achievement gaps among participating countries in PIRLS and TIMSS are shown in Figure 15 and Figure 16, respectively. Figure 15 shows that South Africa had the second-largest pro-girl gap in Grade 4 reading across participating countries. Moreover, South Africa's pro-girl gap is around *four times* larger than the average pro-girl gap out of all countries participating in the 2016 round of PIRLS (49 points compared with 12 points) (Mullis *et al.*, 2017). The magnitude of this gap implies that South African girls are a full year of learning ahead of boys in terms of reading achievement by Grade 4 (Spaull and Makaluza, 2019)[30]. In terms of mathematics achievement, Figure 16 shows that although South Africa's pro-girl achievement gap in TIMSS was much smaller than the pro-girl gap in PIRLS, the magnitude of this pro-girl advantage was the fifth-largest out of the 47 countries participating in TIMSS 2015. Moreover, as the evidence presented in Section 2 makes clear, this pro-girl gap is highly unusual even by regional comparison. These results suggest that girls in South Africa have a unique advantage in primary school reading and mathematics achievement that warrants further investigation.

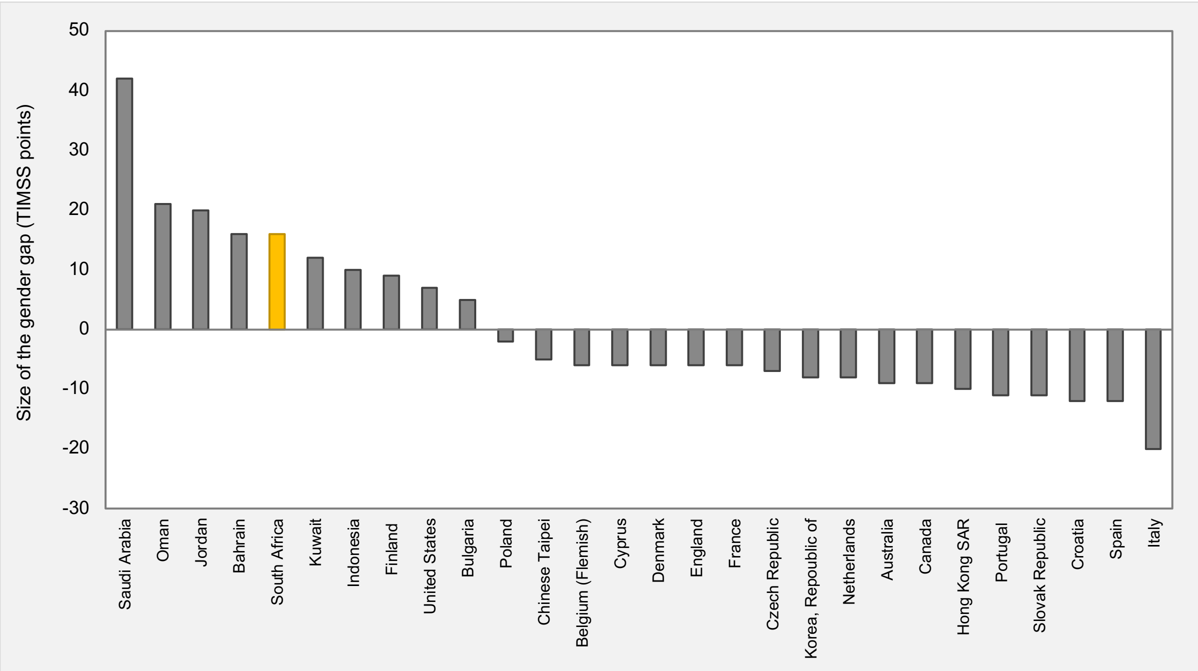
[30] This is an estimate based on Evans and Yuan's (2019) methodology for converting standard deviations in PIRLS scores to equivalent years of schooling, where the increase in test scores between two consecutive grades is assumed to be equal to the amount of learning that takes place in a year. In South Africa, the Grade 5 PIRLS score in 2006 was 0.36 standard deviations higher than the score of Grade 4 learners who wrote the same test (Van der Berg and Gustafsson, 2019), thus 0.36 standard deviations in 2006 PIRLS scores is estimated to be roughly equal to one year of learning in South Africa. Taking into account improvements in the amount of learning that takes place in a year over the period 2011-2016 (Van der Berg and Gustafsson, 2019), half a standard deviation can be considered equivalent to one year of learning in terms of South Africa's PIRLS scores (Spaull and Makaluza, 2019).

Figure 15: Magnitude of the pro-girl gap in PIRLS Reading (2016), by country



Source: Mullis et al. (2016). Notes: Only countries with statistically significant gender gaps in reading are shown. PIRLS scores are standardised to have a standard deviation of 100 points, thus 10 PIRLS points is equal to 10% of a standard deviation.

Figure 16: Magnitude of the pro-girl gap in TIMSS Mathematics (2015), by country



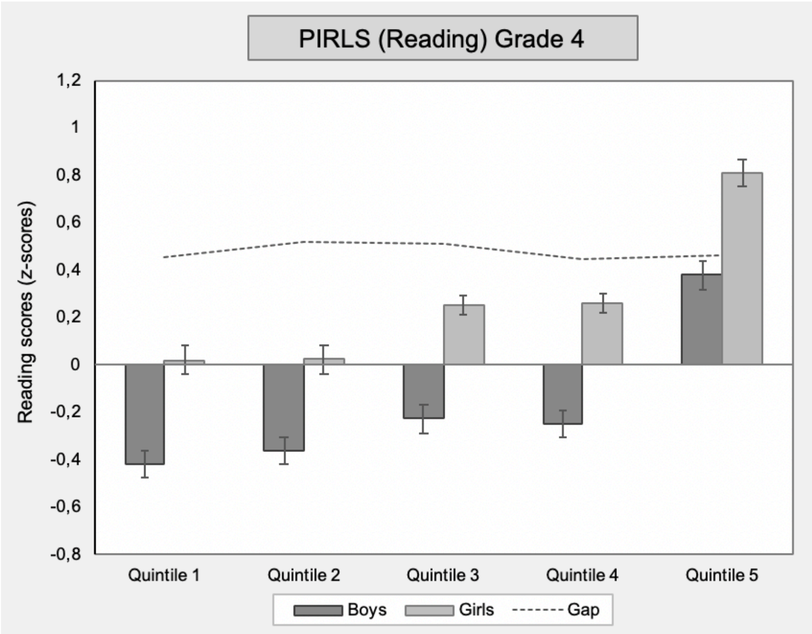
Source: Mullis et al. (2016). Notes: Negative values indicate a pro-boy gap. Only countries with statistically significant gender gaps in mathematics are shown. TIMSS scores are standardised to have a standard deviation of 100 points, thus 10 TIMSS points is equal to 10% of a standard deviation.

Figure 17 and Figure 18 show the magnitudes of South Africa's pro-girl gap (illustrated by the dotted line) by school quintile in PIRLS and TIMSS, respectively. It is clear from Figure 17 that the magnitude of the pro-girl gap in PIRLS is invariant across school quintiles. The absence of SES differences in the magnitude of the pro-girl gap in reading achievement is in itself a noteworthy result, given findings from the international literature that pro-girl gaps in reading achievement are more pronounced among low-SES students (Entwisle, Alexander and Olson, 2007). This result is therefore suggestive of potential differences in the interaction between SES and gender in producing reading outcomes in South Africa as compared to the findings from industrialised countries. This is explored further in the multivariate analysis. By contrast, Figure 18 shows that the magnitude of South Africa's pro-girl gap in TIMSS mathematics achievement decreases with school quintile. Importantly, the size of the pro-girl achievement gap in mathematics decreases to insignificance in Quintile 5 schools. This result is consistent with that of Zuze & Beku (2019), who also find a larger pro-girl gap in Grade 5 TIMSS results in no-fee (Quintile 1-3) schools compared to fee-paying (Quintile 4 and 5) schools.

The last result that is worth emphasising in Figure 17 and Figure 18 is that within gender groups, there are starker inequalities by school wealth in mathematics outcomes compared to reading outcomes. For example, girls in Quintile 5 schools outperformed girls in Quintile 1 schools by 1.4 standard deviations in mathematics (Figure 18), while this gap was 79% of a standard deviation in reading scores (Figure 17). Similarly, boys in Quintile 5 schools achieved mathematics scores 1.5 standard deviations above boys in Quintile 1 schools, while this gap was 80% of a standard deviation in reading test scores. These results suggest that controlling for gender, school SES is more predictive of TIMSS mathematics scores than PIRLS reading scores. That is not to say, however, that the SES differences in reading scores are not themselves worth emphasising. The magnitudes of the gaps in PIRLS reading achievement across school quintiles indicate that controlling for gender, students in Quintile 5 schools are more than a year of learning ahead of students in Quintile 4 schools[31], a result that has received much attention in local education research (see for example Spaull and Makaluza (2019)), and warrants re-emphasising here.

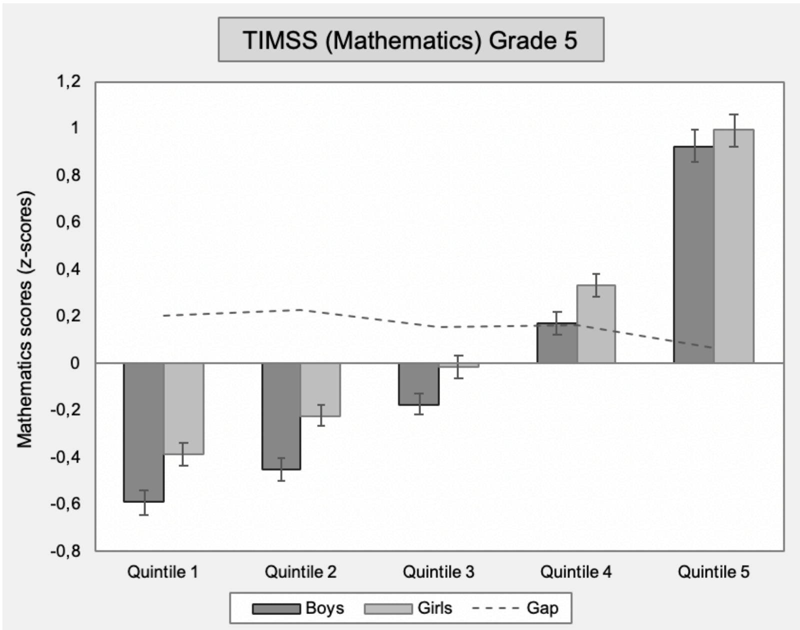
[31] See footnote 30 on page 96.

Figure 17: Magnitude of the pro-girl achievement gap in PIRLS by school quintile



Source: Author's calculations from PIRLS 2016 (Reduced sample of 11,734 students (51% male)). Note: Reading test scores are standardised to have a mean of zero and a standard deviation of one.

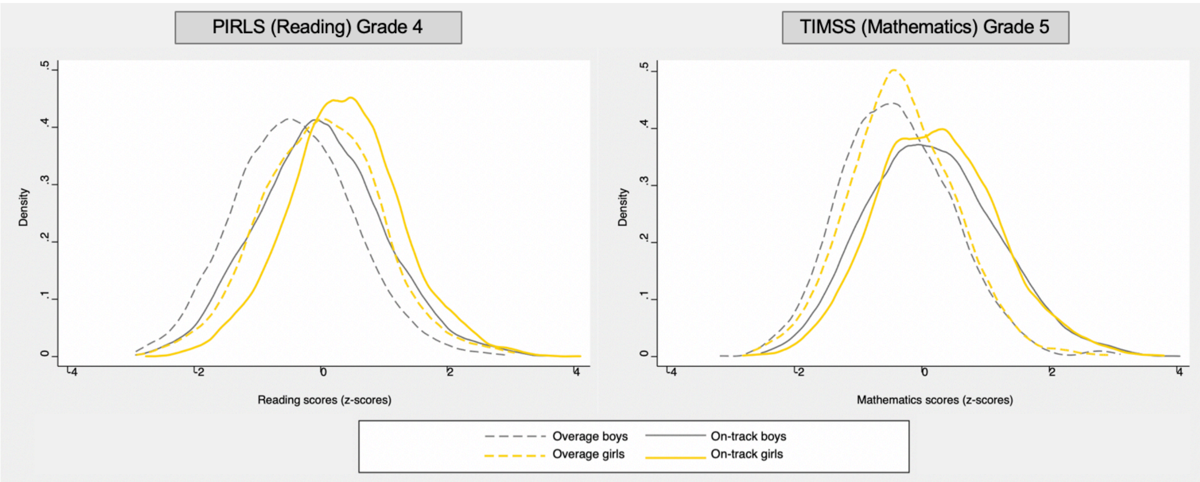
Figure 18: Magnitude of the pro-girl achievement gap in TIMSS by school quintile



Source: Author's calculations from TIMSS Numeracy 2015 (Reduced sample of 10,283 students (52% male)). Note: Mathematics test scores are standardised to have a mean of zero and a standard deviation of one.

The evidence presented in Section 4.3.3 of higher proportions of boys being overage in both PIRLS and TIMSS points to a selection effect whereby any given Grade 4 or 5 class consists of a larger proportion of boys than girls who have repeated a grade. Since grade repetition is associated with weaker academic performance, this implies that any given Grade 4 or 5 class consists of a larger proportion of boys than girls who have been 'selected' to be weaker performers. Given this situation, it is useful to assess whether a pro-girl achievement gap remains after controlling for students' age, i.e. whether there is still a pro-girl gap amongst those who have not repeated. To this end, Figure 19 plots the distributions of test scores in PIRLS and TIMSS for overage and on-track learners, by gender. The figure shows that in TIMSS (the graph on the right-hand side of the figure), the test score distributions of on-track boys and girls (the solid lines) are nearly identical. That is, there is no pro-girl achievement gap in TIMSS when considering only students who are on-track in terms of age. This provides strong evidence in support of the hypothesis that the observed pro-girl achievement gap in TIMSS can be attributed to a 'repetition effect', whereby the fact that boys are more likely to be weaker performers in the early grades translates into a pro-girl advantage in mathematics achievement in Grade 5. By contrast, the PIRLS test score distributions plotted on the left-hand side of Figure 19 reveal a remaining pro-girl achievement gap, even when comparing only students who are on-track in terms of age (the solid lines). This suggests that unlike in TIMSS, girls still outperform boys in PIRLS reading achievement, even when comparing only girls and boys who are in the correct grade-for-age.

Figure 19: PIRLS and TIMSS scores by age and gender



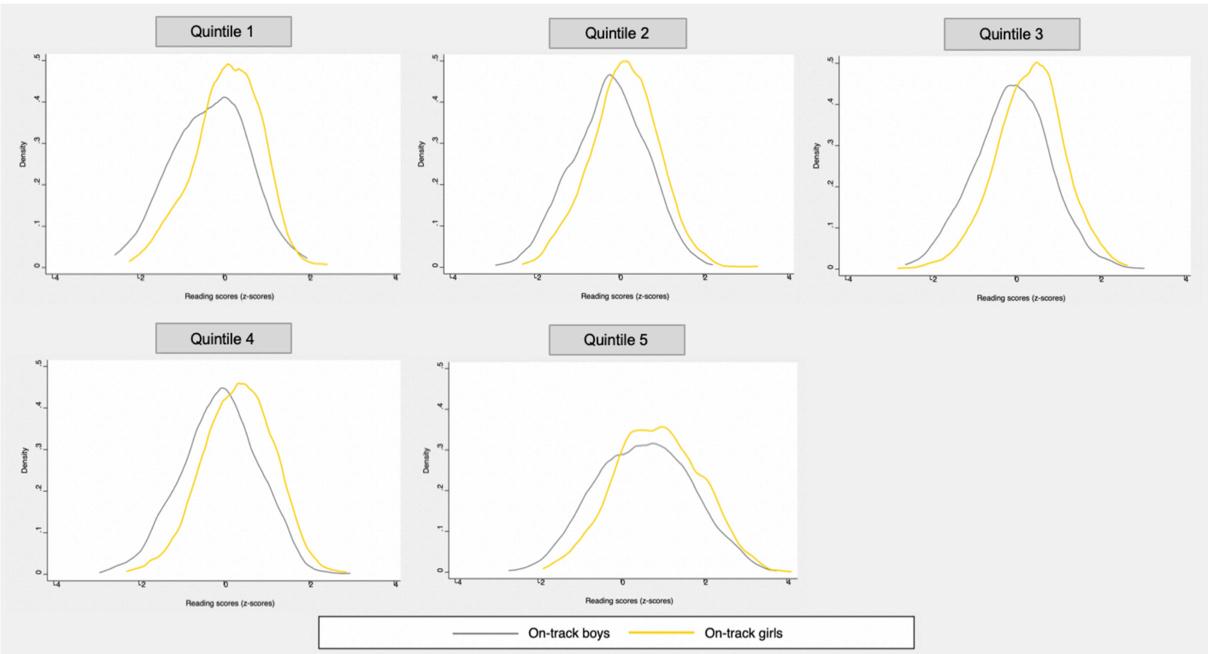
Sources: Author's calculations from PIRLS 2016 (Reduced sample of 11,734 students (51% male)) and TIMSS Numeracy 2015 (Reduced sample of 10,283 students (52% male)). Reading and mathematics test scores are standardised to have a mean of zero and a standard deviation of one.

Given the evidence from Table 14 and Table 15 that the proportion of overage students differs across school quintiles, the achievement distributions plotted in Figure 19 may be masking important variation in the relationship between students' age and achievement in PIRLS and TIMSS by school quintile. To determine whether the achievement distributions plotted in Figure 19 hold across school quintiles, PIRLS and TIMSS test score distributions by gender and school quintile are plotted in Figure 20 and Figure 21, respectively, this time only for students who are on-track in terms of age. Figure 20 shows that the magnitude pro-girl achievement gap among on-track students in PIRLS differs by school quintile, with a more pronounced gap observed in poorer schools. It is interesting to note that despite this variation, a pro-girl gap is observable in all school quintiles, even when restricting the sample to students who are on-track in terms of age. This constitutes evidence that the result presented in Figure

19 of a large remaining pro-girl gap in PIRLS reading achievement even among on-track students holds across all school quintiles.

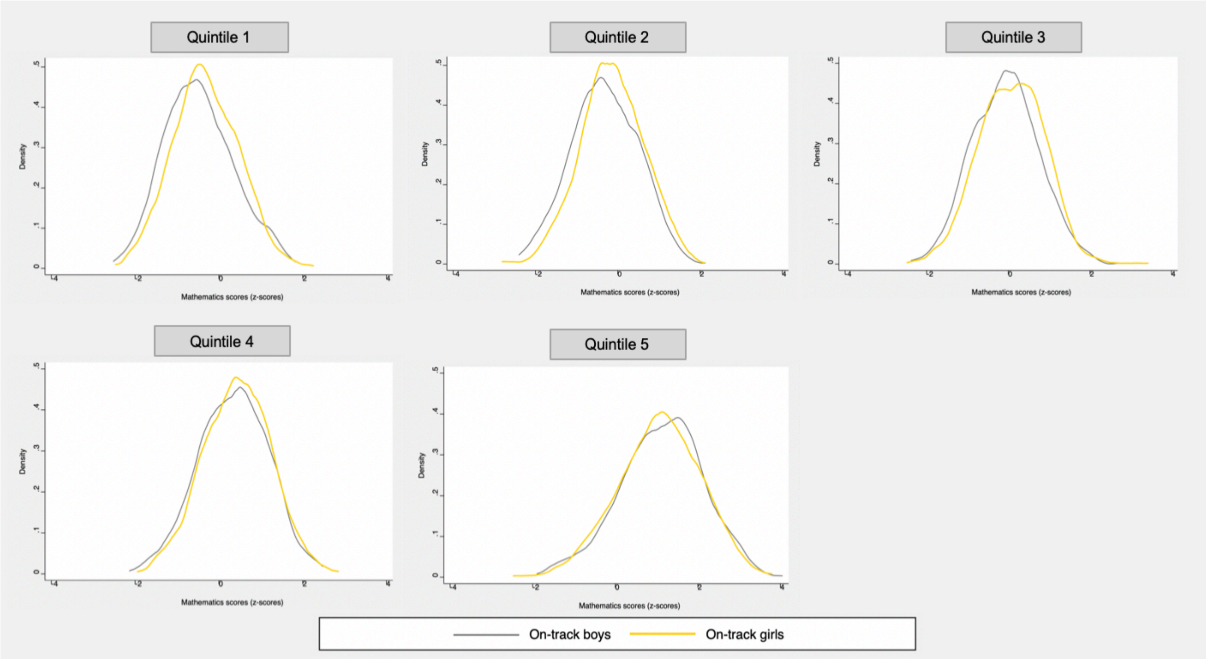
Interestingly, Figure 21 points to more variation across school quintiles in the pro-girl achievement gap among on-track students in TIMSS than is the case for PIRLS, with girls' Grade 5 mathematics test score distributions lying slightly to the right of boys' in Quintiles 1-3, and Quintile 4 and 5 boys' and girls' distributions being almost identical. This result, in combination with the results presented in Figure 20, suggests there may be SES differences in the relative contribution of students' age to the observed pro-girl achievement gaps in both PIRLS and TIMSS. Specifically, the fact that a large pro-girl gap is observable among on-track students in poorer schools, but that there seems to be almost no gender gap among on-track students in Quintile 4 and 5 students in TIMSS, suggests that the contribution of the 'overage' variable to the observed pro-girl achievement gap is relatively *larger* in wealthier schools. This is an unexpected result, since one would expect the contribution of being overage to the pro-girl achievement gap to be larger in poorer schools where grade repetition is more common. This result is explored further in the decomposition analysis that follows.

Figure 20: Distribution of PIRLS scores by gender and school quintile (on-track students)



Sources: Own calculations from PIRLS 2016. Notes: Reading test scores are standardised to have a mean of zero and a standard deviation of one. The test score distributions of on-track students are statistically significantly different from overage students at p<0.01 in all school quintiles (according to Kolmogorov-Smirnov tests of equality of distribution functions).

Figure 21: Distribution of TIMSS scores by gender and school quintile (on-track students)



Sources: Own calculations from TIMSS Numeracy 2015. Notes: Mathematics test scores are standardised to have a mean of zero and a standard deviation of one. The test score distributions of on-track students are statistically significantly different from overage students at p<0.01 in Quintiles 1, 2, and 3 (according to Kolmogorov-Smirnov tests of equality of distribution functions).

4.5. THE SOURCES OF THE GENDER ACHIEVEMENT GAP

4.5.1. THE DECOMPOSITION APPROACH

Decomposition analysis has long been used by labour economists to identify the relative importance of various factors that contribute to gender, race or other gaps in labour market outcomes (Cobb-Clark and Moschion, 2017). More recently, a number of economists of education have used this approach to examine disparities in learning outcomes by, for example, urban-rural status (Burger, 2011), as well as gender (Kingdon, 2002; Cobb-Clark and Moschion (2017). In essence, this approach allows one to separate gender gaps into two components: (i) The component that can be explained due to differences in observable characteristics of males and females; and (ii) The unexplained component. My aim is to decompose the pro-girl gaps in achievement in both PIRLS and TIMSS into these two components. Formally, assuming a linear model of achievement, boys' and girls' test scores can be expressed as

$$\bar{T}_G^j = \bar{X}_G \hat{\beta}_G^j + \bar{\varepsilon}_G, \qquad \bar{\varepsilon}_G = 0; G \in (M, F) \qquad (1)$$

where $\bar{T}_G^j$ denotes the mean test score of students of gender G (male (M) or female (F)) in subject j, $\bar{X}_G$ denotes the mean endowments (observable characteristics) of students of that gender, $\hat{\beta}_G^j$ denotes the coefficients on those endowments (that is, how those endowments are translated into test scores) for each gender and in each subject, and $\bar{\varepsilon}_G$ denotes the error term, which we assume to be zero. The gender gap in test scores can therefore be expressed as

$$\bar{T}_M^j - \bar{T}_F^j = \bar{X}_M \hat{\beta}_M^j - \bar{X}_F \hat{\beta}_F^j \tag{2}$$

I adopt Cobb-Clark and Moschion's (2017) approach of introducing a gender neutral coefficient vector ($\hat{\beta}_P^j$) to determine the contribution of the gender differences in endowments such that

$$\bar{X}_M \hat{\beta}_M^j - \bar{X}_F \hat{\beta}_F^j = (\bar{X}_M^j - \bar{X}_F^j)\, \hat{\beta}_P^j + \{\bar{X}_M^j (\hat{\beta}_M^j - \hat{\beta}_P^j) + \bar{X}_F^j (\hat{\beta}_P^j - \hat{\beta}_F^j)\} \tag{3}$$

where $\hat{\beta}_P^j$ is the coefficient from a pooled ordinary least squares regression of test scores on the full set of covariates over both males and females, and $\hat{\beta}_M^j$ and $\hat{\beta}_F^j$ are coefficients from separate regressions for males and females, respectively (Jann, 2008b). Thus the first term on the right-hand side of equation (3) is the explained component of the gender gap in test scores, that is, the difference in boys' and girls' test scores that arises because boys and girls have different endowments of the characteristics that matter for achievement (the 'endowment effect'). These characteristics are evaluated (i.e. weighted) using the vector of gender-neutral responses ($\hat{\beta}_P^j$) (Jann, 2008). The second term on the right-hand side of equation (3) is the unexplained component of the gender gap (the 'response effect'). This term can be interpreted as the part of the gender gap that arises because boys' and girls' endowments are not translated into test scores in a gender-neutral way (Cobb-Clark and Moschion, 2017). However, given that this interpretation requires the strong assumption that all factors that matter for achievement are included in the model (Jann, 2008), the coefficients on the covariates contributing to the unexplained component should be interpreted with caution.

To account for the potential sample bias resulting from different proportions of boys and girls being overage in both PIRLS and TIMSS, I conduct two decomposition analyses: One using each full PIRLS and TIMSS sample, and one using only the students in each dataset that are in the appropriate grade for their age. Conducting two decomposition analyses allows me to first assess the contribution of being overage (that is, having repeated at least a year) to the observed pro-girl achievement gaps in PIRLS and TIMSS, and then to investigate whether there are differences in observable characteristics between boys and girls contributing to the pro-girl achievement gap that remain when considering only "on-track" students in both datasets.

4.5.2. SOURCES OF THE PRO-GIRL ACHIEVEMENT GAP

The results from the set of first decompositions are presented in Table 16 and Table 17. Since the aim of the first decomposition is to determine how much of the pro-girl gap in both datasets can be explained by gender differences in the proportion of overage students, only these results are reported in the tables[32]. The tables show boys' and girls' average standardised PIRLS and TIMSS scores and the magnitude of the gender gap in test scores, by school quintile. Negative values indicate a pro-girl

[32] The detailed results from the first set of decompositions are reported in Table C14 and Table C15 of Appendix C.

advantage. The share of the gender gap in test scores attributable to differences in boys' and girls' characteristics (the 'endowment effect') is reported in the "Explained" row. Differences between boys and girls in how endowments are translated into achievement (the 'response effect') are reported in the "Unexplained" row. Figure 22 and Figure 23 show the results from Table 16 and Table 17 graphically. The dark grey bars represent the size of the endowment effect as a proportion of the total gender gap (the dotted lines in the figures), while the light grey bars represent the proportion of the gender gap that is explained by the dummy variable indicating whether students are overage[33]. For example, the explained component of the pro-girl gap in Quintile 1 schools in PIRLS is equal to 0.135, which constitutes 27% of the total gender gap (0.503) in Quintile 1 schools – the first dark grey bar in Figure 22. The coefficient on the "overage" dummy among Quintile 1 students is 0.036, which is 7% of the total gender gap – the first light grey bar in Figure 22.

Table 16: Selected results from the first decomposition (Includes overage students): PIRLS (Grade 4)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Gender gap	-0.503***	-0.385***	-0.493***	-0.556***	-0.397***
	(0.051)	(0.027)	(0.031)	(0.029)	(0.041)
Boys' average	-0.457***	-0.415***	-0.234***	-0.274***	0.586***
	(0.062)	(0.045)	(0.047)	(0.050)	(0.100)
Girls' average	0.046	-0.030	0.259***	0.282***	0.984***
	(0.054)	(0.029)	(0.034)	(0.043)	(0.089)
Explained	-0.135***	-0.161***	-0.196***	-0.200***	-0.168***
	(0.025)	(0.014)	(0.022)	(0.022)	(0.031)
Unexplained	-0.368***	-0.224***	-0.296***	-0.356***	-0.230***
	(0.043)	(0.029)	(0.027)	(0.032)	(0.031)
Overage	-0.036***	-0.040***	-0.047***	-0.038***	-0.031***
	(0.010)	(0.008)	(0.009)	(0.006)	(0.008)
N	2,024	2,256	2,560	2,503	2,391

Notes: The decompositions include all the controls described in Section 0, but are not reported here. See Table B15 of Appendix B for the full list of controls and the detailed results. Standard errors are calculated at the school level and reported in parentheses.
*** p < 0.01; ** p < 0.05; * p < 0.1.

[33] Only statistically significant components are plotted. The explained component and the contribution of 'overage' to the gender gap in quintile 5 schools in TIMSS are not plotted since there is no statistically significant gender achievement gap among these students.

Table 17: Selected results from the first decomposition (Includes overage students): TIMSS (Grade 5)

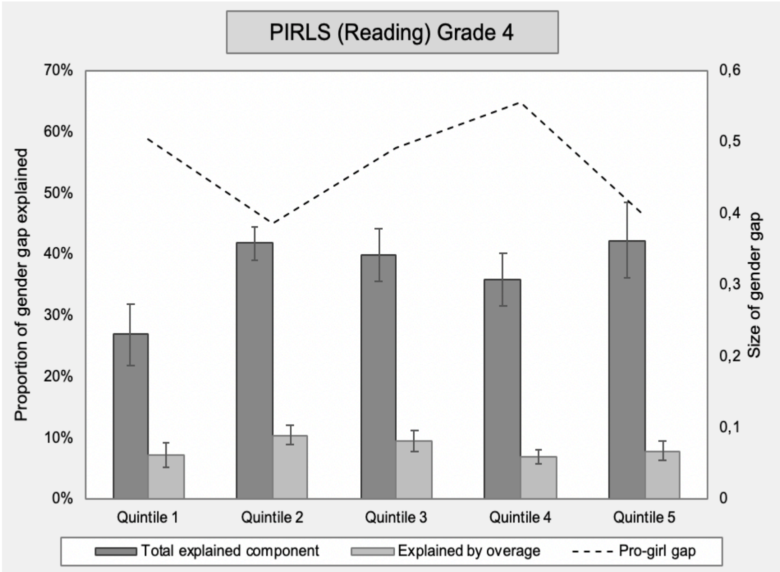
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Gender gap	-0.216***	-0.209***	-0.223***	-0.156***	-0.010
	(0.062)	(0.032)	(0.026)	(0.026)	(0.073)
Boys' average	-0.611***	-0.517***	-0.281***	0.177***	0.956***
	(0.081)	(0.040)	(0.030)	(0.044)	(0.081)
Girls' average	-0.395***	-0.309***	-0.058	0.333***	0.966***
	(0.053)	(0.043)	(0.031)	(0.040)	(0.102)
Explained	-0.044**	-0.120***	-0.072***	-0.089***	-0.087
	(0.018)	(0.020)	(0.014)	(0.022)	(0.053)
Unexplained	-0.172**	-0.088***	-0.151***	-0.066***	0.077
	(0.065)	(0.025)	(0.027)	(0.014)	(0.054)
Overage	-0.014	-0.057***	-0.031***	-0.037***	-0.053***
	(0.008)	(0.008)	(0.006)	(0.008)	(0.015)
N	1,802	2,084	2,207	2,280	1,774

Notes: The decompositions include all the controls described in Section 0, but are not reported here. See Table C16 of Appendix C for the full list of controls and the detailed results. Standard errors are calculated at the school level and reported in parentheses.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

The results presented in Figure 22 and Figure 23 indicate that on average across school quintiles, around 40% of the pro-girl gap in PIRLS can be explained by gender differences in the observable characteristics included in the decomposition model, while this proportion is 50% for the pro-girl gap in TIMSS. Importantly, the unexplained component of the gender gap (the 'response effect') is larger than the explained component (in all school quintiles in PIRLS, and in Quintiles 1 and 3 in TIMSS). In other words, girls have an advantage both in terms of their endowments of the characteristics considered here, as well as in terms of how those endowments are transformed into achievement, with the latter effect dominating the former in most cases. Since the focus of the first set of decompositions is the contribution of the 'overage' variable to the explained component of the gender gap, I return to consideration of the relative sizes of the explained and unexplained components in my discussion of the results of the second set of decompositions.

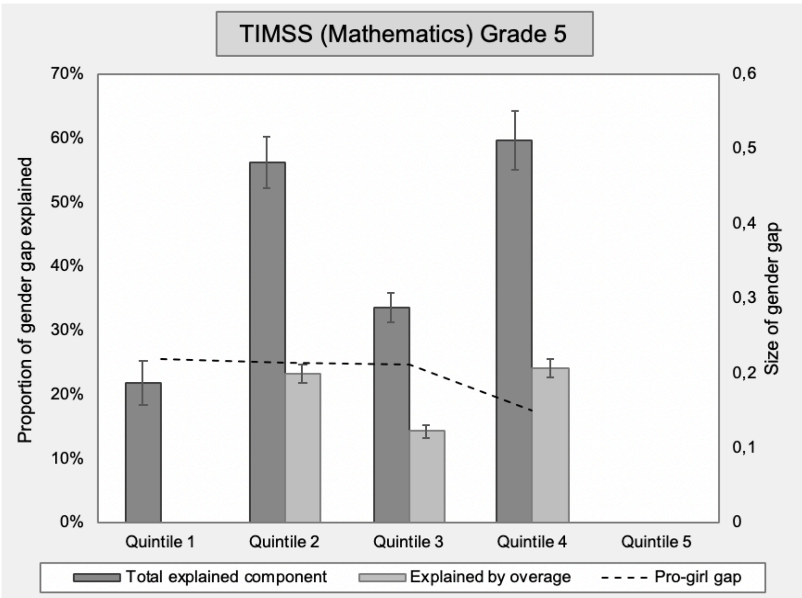
The results from the first decomposition further confirm the descriptive results presented in Section 4.3.3 of the importance of gendered patterns in grade repetition in the early grades in accounting for South Africa's pro-girl advantage in Grade 4 reading and Grade 5 mathematics achievement: The 'overage' variable emerges as significantly contributing to the pro-girl achievement gap across virtually all school quintiles in both PIRLS and TIMSS. Interestingly, however, it is clear from Figure 22 and Figure 23 that the contribution of the overage variable to the explained components of the gender gaps in both datasets is relatively small – around 10% in PIRLS and roughly 20% in TIMSS. This suggests that while gender differences in the proportions of students who are overage do explain part of the pro-girl achievement gaps in both PIRLS and TIMSS, a large part of this gap can be explained by differences between boys and girls in other observable characteristics. This result is explored further in the second set of decompositions.

Figure 22: Magnitude of the explained component of the gender gap in PIRLS, by school quintile



Source: Author's calculations from PIRLS 2016 (Reduced sample of 11,734 students (51% male)).

Figure 23: Magnitude of the explained component of the gender gap in TIMSS, by school quintile



Source: Author's calculations from TIMSS Numeracy 2015 (Reduced sample of 10,283 students (52% male)). Notes: No "overage" component plotted for Quintile 1 schools since the overage variable does not significantly contribute to the explained component of the gender gap in these schools. No components plotted for Quintile 5 schools since the gender achievement gap among students in these schools is not statistically significant.

The results of the second set of decompositions are presented in Table 18 and Table 19. The top panels show boys' and girls' average standardised test scores and the magnitude of the gender gap in test scores, by school quintile. Negative values indicate a pro-girl advantage. The share of the gender gap in test scores attributable to differences in boys' and girls' characteristics (differences in endowments) are presented in the "Explained" columns of the tables. The share of the gender gap attributable to differences between boys and girls in the way endowments are translated into achievement are reported in the "Unexplained" columns. For example, the negative and significant coefficients on the

"Confidence index" across all school quintiles in PIRLS (Table 18) indicate that girls have higher confidence in reading than boys, and that this difference in confidence contributes significantly to the gender gap in reading scores in all quintiles. The lack of a significant response effect in student confidence across quintiles indicates that there are no gender differences in how given endowments of confidence are translated into PIRLS test scores. The magnitudes of the explained components are presented graphically in Figure 24 and Figure 25, where the yellow bars represent the explained component (endowment effect) as a proportion of the total gender gap (the dotted line). The grey bars in Figure 24 represent the magnitudes of the pro-girl advantage in endowments of student attitudes, expressed as a proportion of the gender gap. Differences in endowments of student attitudes are not plotted in Figure 25, since these differences are not statistically significant (see Table 19).

The first noteworthy result from the second set of decompositions is that a statistically significant pro-girl advantage remains across all school quintiles in PIRLS, and Quintiles 1-4 in TIMSS, even when accounting for gendered repetition patterns by limiting the sample to only students that are in the appropriate grade for their age. While the magnitude of the gender gap decreases in almost all school quintiles in PIRLS (with the exception of Quintile 1), the remaining pro-girl achievement gap remains large, suggesting that girls' advantage in achievement in the Foundation Phase and consequently lower overage proportion only partially accounts for South Africa's unusually large pro-girl achievement gap in PIRLS. A similar result emerges in the TIMSS data, whereby restricting the sample to exclude overage students reduces the magnitude of the gender in Quintiles 1-4, but a significant pro-girl achievement gap remains.

The unshaded bars in Figure 25 indicate that the explained component of the pro-girl gap in TIMSS mathematics scores is not statistically significant in either Quintile 1 or Quintile 4, a noteworthy result. This suggests that none of the observable characteristics included in the model contribute meaningfully to the observed pro-girl achievement gap in Quintiles 1 and 4 in TIMSS – that is, boys and girls in these school quintiles do not differ in their endowments of the characteristics that matter for mathematics achievement. By contrast, the explained component of the gender gap in PIRLS remains significant across all school quintiles, even after limiting the sample to on-track students. Importantly, the size of the endowment effect increases with school socio-economic quintile, from 19% in Quintile 1 to around 30% in Quintiles 2-5 (Figure 24). This points to SES differences in the nature of the pro-girl achievement gap that remain after restricting the sample to only students who are on-track in terms of age, with the observable characteristics considered explaining a larger proportion of the gender gap in wealthier schools.

The unexplained components are statistically significant in all school quintiles with a significant gender gap in both PIRLS and TIMSS. This is a noteworthy result, since it suggests that there is something about the way endowments are translated into achievement that differs between boys and girls in both

PIRLS and TIMSS. The results from the second set of decompositions therefore suggest that whereas girls have both more endowments *and* higher returns to those endowments in PIRLS, the pro-girl gap in TIMSS is mostly driven by girls' higher returns to endowments. Given the well-cited issues with interpreting individual coefficients on the unexplained component (Jann, 2008), however, we will refrain from placing too much emphasis on these coefficients. The main result here is that the second set of decompositions constitute evidence of gender differences in both endowments and the returns to endowments in PIRLS, whereas the remaining pro-girl gap in TIMSS is largely due to gender differences in returns to endowments.

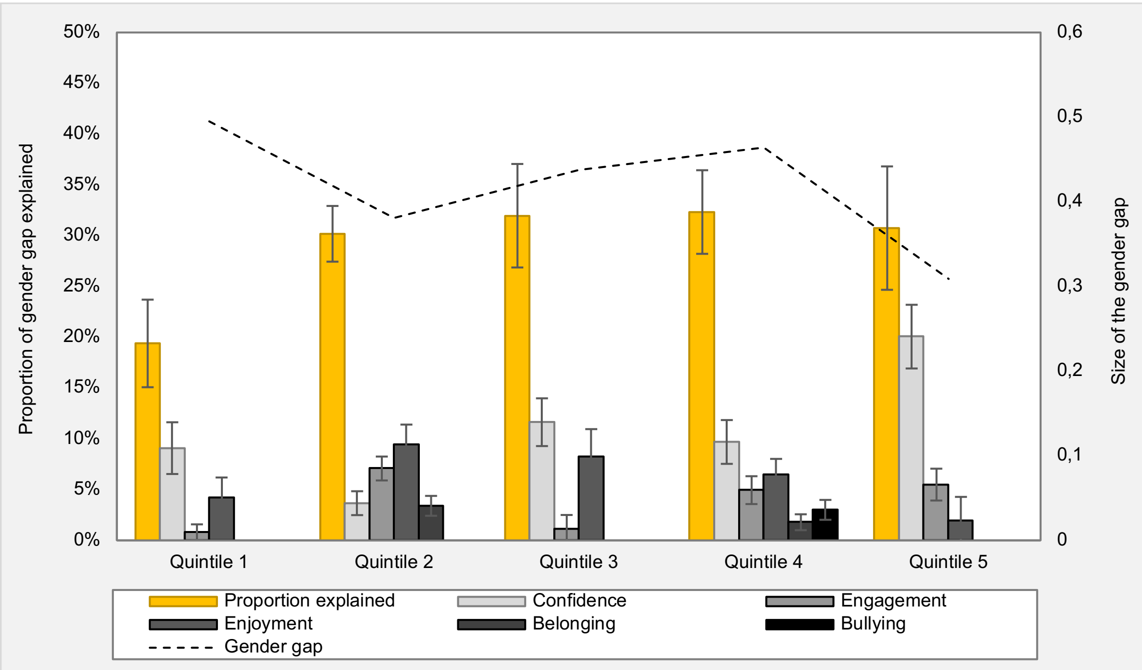
In terms of gender differences in student attitudes, the results presented in Figure 24 suggest that girls' higher endowments of these characteristics contribute significantly to the pro-girl achievement gap, even when restricting the sample of students to only those who are on-track in terms of age. The same is not true for TIMSS (Table 19), where even in the two quintiles where the explained component of the gender gap is statistically significant (Quintiles 2 and 3), none of the subject-specific student attitudes contribute consistently to the gender achievement gap in both quintiles.

It is further interesting to note that in addition to SES differences in the magnitude of the gender gap, the results in Figure 24 point to significant variation across school quintiles in the contribution of student attitudes to the pro-girl achievement gap in PIRLS. For example, girls' higher reading self-concept explains around 10% of the gender achievement gap in Quintile 1-4 schools, and twice as much (20%) in Quintile 5 schools. This is a noteworthy result, since it points to variation in the association between student attitudes and learning outcomes by school socio-economic context. The interaction between student attitudes and SES has received little attention in both the local and the international literature, and the findings presented here suggest it is likely to be a fruitful avenue for future attempts to understand the role of these factors in determining learning outcomes, especially in developing country contexts.

In addition to attitudes towards reading and school, girls also reported doing homework more frequently than boys in all five school quintiles in PIRLS, as well as in the two quintiles with a significant explained component in TIMSS (Quintiles 2 and 3). Interestingly, the fact that girls report doing homework more often than boys accounts for a larger proportion of the explained component of the pro-girl gap in reading among Quintile 5 students (11%) than is the case for students in poorer school quintiles (for example, 4% among students in Quintile 1 schools). The result that girls report doing homework more frequently and that this contributes to the pro-girl achievement gap across virtually all school quintiles constitutes some evidence in support of the hypothesis that part of the pro-girl advantage in educational outcomes is due to girls exhibiting more of the behaviours that are rewarded in school and support learning. Moreover, the fact that this variable contributes more to the gender gap among students in wealthier schools in PIRLS suggests the return on doing homework - in terms of learning outcomes - is higher in better-resourced schools than in more disadvantaged schools.

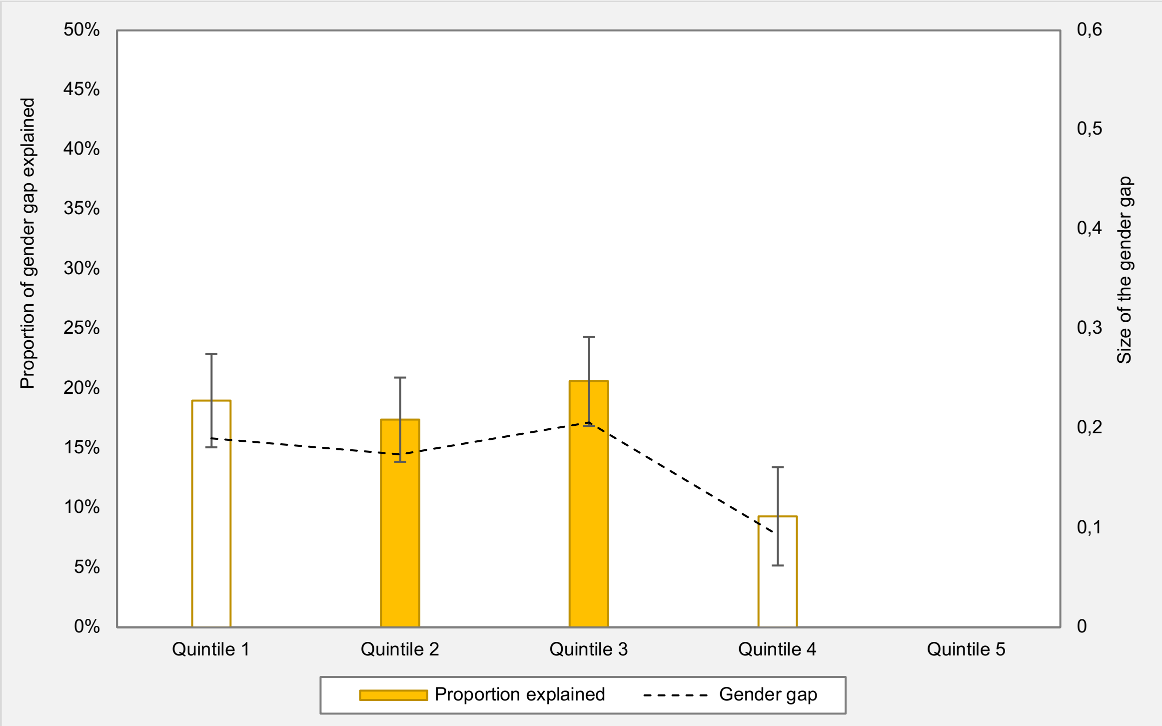
It is interesting to note that the pro-boy advantage in asset index scores observed in Section 4.3.3 does not hold in a multivariate context: The lack of statistically significant coefficients on the asset index (reported in the "Explained" column) in Table 18 and Table 19 suggests that boys and girls do not differ significantly in terms of their endowments of home assets in either PIRLS or TIMSS. This result constitutes some evidence that boys were more prone to over-reporting on the home possession questionnaire items, that is, that there are not in actual fact significant differences in student wealth by gender.

Figure 24: Magnitude of explained component of the gender gap in PIRLS, by school quintile, for samples that are the correct age-for-grade



Sources: Author's calculations from PIRLS 2016 (Restricted sample of 8,022 students (46% male)). Note: Only student attitudes that significantly contribute to the gender gap in achievement are plotted.

Figure 25: Magnitude of explained component of the gender gap in TIMSS, by school quintile, for samples that are the correct age-for-grade



Source: Author's calculations from TIMSS Numeracy 2015 (Restricted sample of 7,409 students (46% male)). Notes: Unshaded bars indicate that the explained component is not statistically significant. No components plotted for Quintile 5 schools since the gender achievement gap among students in these schools is not statistically significant. No components plotted for student attitudes since these variables do not significantly contribute to the explained component of the gender gap.

Table 18: Results from the second decomposition (only students who are the correct age-for-grade): PIRLS (Grade 4)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5	
Gender gap	-0.495*** (0.051)		-0.381*** (0.032)		-0.438*** (0.036)		-0.464*** (0.038)		-0.309*** (0.048)	
Boys' average	-0.376*** (0.085)		-0.308*** (0.050)		-0.084 (0.047)		-0.084 (0.047)		0.712*** (0.105)	
Girls' average	0.119 (0.061)		0.073*** (0.032)		0.354*** (0.040)		0.354*** (0.040)		1.021*** (0.094)	
	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained
Confidence index	-0.045*** (0.013)	-0.026 (0.013)	-0.014*** (0.006)	-0.000 (0.001)	-0.051*** (0.012)	0.001 (0.005)	-0.045*** (0.011)	-0.004 (0.003)	-0.062*** (0.016)	-0.003 (0.015)
Engagement index	-0.004 (0.004)	-0.004 (0.005)	-0.027*** (0.006)	-0.000 (0.001)	-0.005 (0.007)	-0.002 (0.003)	-0.023*** (0.007)	0.000 (0.005)	-0.017** (0.008)	-0.003 (0.013)
Enjoyment index	-0.021 (0.010)	0.002 (0.004)	-0.027** (0.010)	0.000 (0.001)	-0.036** (0.014)	-0.013 (0.009)	-0.030*** (0.008)	-0.006 (0.007)	-0.006 (0.012)	0.003 (0.004)
Belonging index	-0.000 (0.002)	-0.012 (0.013)	-0.013** (0.005)	-0.001 (0.004)	-0.020 (0.013)	-0.004 (0.005)	-0.010** (0.004)	0.001 (0.003)	-0.003 (0.003)	-0.000 (0.001)
Bullying index	-0.014 (0.009)	0.003 (0.010)	-0.016 (0.007)	0.005 (0.003)	-0.011 (0.005)	-0.001 (0.008)	-0.014** (0.005)	0.001 (0.002)	-0.010 (0.006)	0.005 (0.009)
Homework	-0.018** (0.008)	0.015 (0.018)	-0.023*** (0.005)	-0.035 (0.026)	-0.013** (0.005)	0.012 (0.024)	-0.020*** (0.006)	-0.008 (0.017)	-0.033** (0.013)	0.022 (0.044)
Attended ECD	-0.005 (0.003)	-0.017 (0.124)	-0.002 (0.001)	-0.165 (0.105)	0.001 (0.002)	-0.171 (0.126)	-0.001 (0.001)	-0.292*** (0.093)	0.006 (0.003)	-0.036 (0.149)
Asset index	-0.001 (0.005)	-0.094** (0.038)	-0.000 (0.001)	-0.026** (0.012)	-0.005 (0.004)	0.007 (0.004)	0.003 (0.002)	0.001 (0.010)	0.001 (0.002)	-0.055 (0.028)
First language	0.016** (0.007)	-0.295*** (0.056)	0.005** (0.002)	-0.018 (0.072)	-0.001 (0.001)	0.217*** (0.064)	-0.003 (0.002)	-0.014 (0.060)	0.007 (0.005)	-0.000 (0.054)
African language school	-0.007** (0.003)	0.077 (0.126)	-0.012** (0.005)	0.147 (0.072)	0.007 (0.005)	-0.087** (0.040)	-0.009 (0.007)	0.050 (0.048)	0.020 (0.018)	-0.010 (0.022)
School has a library	-0.001 (0.001)	-0.002 (0.022)	-0.000 (0.001)	0.001 (0.015)	0.001 (0.003)	-0.031 (0.019)	0.002 (0.004)	0.080** (0.037)	0.000 (0.002)	0.021 (0.036)
School has computers	-0.003 (0.004)	-0.031 (0.019))	-0.000 (0.002)	0.022 (0.014)	0.001 (0.002)	0.024 (0.011)	0.000 (0.003)	-0.008 (0.047)	0.000 (0.001)	0.091** (0.034)
Constant		-0.006 (0.160)		-0.230 (0.140)		-0.275 (0.181)		-0.166 (0.085)		-0.460*** (0.153)
Total	-0.096*** (0.022)	-0.400*** (0.041)	-0.115*** (0.014)	-0.266*** (0.030)	-0.140*** (0.026)	-0.299*** (0.025)	-0.150*** (0.021)	-0.314*** (0.038)	-0.095*** (0.031)	-0.214*** (0.034)
N	1,347		1,514		1,698		1,707		1,756	

Notes: All models include controls for province. Standard errors are calculated at the school level and reported in parentheses. *** p < 0.01; ** p < 0.05; * p < 0.1.

Table 19: Results from the second decomposition (only students who are the correct age-for-grade): TIMSS (Grade 5)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5	
Gender gap	-0.190***		-0.174***		-0.206***		-0.093***		0.134	
	(0.050)		(0.041)		(0.026)		(0.031)		(0.083)	
Boys' average	-0.532***		-0.398***		-0.196***		0.304***		1.180***	
	(0.088)		(0.040)		(0.031)		(0.042)		(0.070)	
Girls' average	-0.342***		-0.224***		0.010		0.397***		1.046***	
	(0.058)		(0.047)		(0.034)		(0.040)		(0.097)	
	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained
Confidence index	-0.004	-0.016	-0.004	-0.000	-0.005	0.000	0.014	0.003	0.049***	0.027
	(0.008)	(0.012)	(0.008)	(0.003)	(0.006)	(0.001)	(0.009)	(0.004)	(0.016)	(0.014)
Engagement index	0.012	-0.000	0.012	0.023**	-0.002	0.004	-0.000	-0.013	0.000	-0.018
	(0.009)	(0.006)	(0.009)	(0.010)	(0.002)	(0.004)	(0.003)	(0.009)	(0.001)	(0.010)
Enjoyment index	-0.018	0.006	-0.018	0.002	-0.005	-0.002	0.000	0.015	0.003	-0.006
	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.002)	(0.008)	(0.011)	(0.008)	(0.008)
Belonging index	-0.013	0.003	-0.013	-0.006	-0.003	-0.008	0.000	-0.003	-0.001	0.009
	(0.006)	(0.002)	(0.006)	(0.005)	(0.004)	(0.004)	(0.001)	(0.005)	(0.005)	(0.008)
Bullying index	0.001	-0.008	0.001	-0.009	-0.017***	-0.001	-0.019**	-0.001	-0.007	-0.029
	(0.004)	(0.006)	(0.004)	(0.006)	(0.004)	(0.001)	(0.007)	(0.003)	(0.006)	(0.017)
Homework	0.001	-0.126**	0.001	0.075**	-0.007**	0.032	-0.023***	-0.007	-0.006	0.095
	(0.002)	(0.046)	(0.002)	(0.035)	(0.003)	(0.033)	(0.006)	(0.026)	(0.004)	(0.054)
Attended ECD	-0.001	0.032	-0.001	-0.041	-0.001	0.071**	-0.000	0.049	0.001	-0.019
	(0.004)	(0.079)	(0.004)	(0.053)	(0.001)	(0.028)	(0.001)	(0.042)	(0.002)	(0.043)
Asset index	-0.004	0.048	-0.004	-0.011**	0.001	-0.001	-0.003	-0.014	0.001	-0.054
	(0.008)	(0.031)	(0.008)	(0.004)	(0.002)	(0.002)	(0.003)	(0.021)	(0.005)	(0.055)
First language	-0.002	0.006	-0.002	-0.036***	-0.002	0.007	-0.014	-0.027	-0.020	-0.049
	(0.004)	(0.016)	(0.004)	(0.011)	(0.002)	(0.011)	(0.009)	(0.026)	(0.018)	(0.045)
School has a library	-0.004	0.020	-0.002	0.040**	0.003	0.028	-0.001	-0.040	0.020	0.126
	(0.005)	(0.014)	(0.003)	(0.017)	(0.002)	(0.015)	(0.001)	(0.026)	(0.012)	(0.187)
School has computers	0.000	0.024	-0.001	-0.009	0.001	-0.013	0.005	-0.059**	0.043	-0.034
	(0.001)	(0.013)	(0.004)	(0.013)	(0.003)	(0.022)	(0.004)	(0.025)	(0.028)	(0.154)
Constant		-0.150		-0.132		-0.284***		0.046		0.014
		(0.119)		(0.074)		(0.037)		(0.071)		(0.183)
Total	-0.033	-0.157***	-0.070***	-0.105**	-0.040**	-0.166***	-0.041	-0.052**	0.072	0.062
	(0.020)	(0.048)	(0.018)	(0.037)	(0.019)	(0.019)	(0.021)	(0.019)	(0.055)	(0.060)
N	1,178		1,531		1,527		1,718		1,455	

Notes: All models include controls for province. Standard errors are calculated at the school level and reported in parentheses. *** p < 0.01; ** p < 0.05; * p < 0.1.

4.6. ROBUSTNESS CHECKS

The robustness of these results was evaluated against (i) the inclusion of additional classroom -level covariates; (ii) a different calculation of the student asset indices used to split the sample into school quintiles; and (iii) different ways of dealing with missing data in the student attitude measures. In terms of the first robustness check, teacher characteristics (gender, age, and years of experience) were included in the decomposition model. The results from this model specification are reported in Table C16 and Table C17 of Appendix C. The tables show that the main results of the second decomposition (reported in Table 18 and Table 19) are robust to the inclusion of additional classroom-level covariates (specifically teacher characteristics). To perform the second robustness check, I constructed an SES index that included parental education and split the samples into school quintiles based on this SES index. To do so, missing values on the parent education variables were imputed as the modal parental education at the school level. Schools that had missing values for all students on the parental education variable were assigned the lowest category (primary schooling). The results from these estimations are presented in Table C18 of Appendix C. The results from the main decomposition are robust to this alternative way of constructing the SES index. Lastly, following Shepherd (2013), missing data on the student attitude variables were assigned the lowest value the variable could take on and missing values on the parent education variables were imputed as the lowest category (some primary schooling). The results from these estimations are presented in Table C19 of Appendix C. The main results presented in Section 4.5 are also robust this alternative way of dealing with missing data in the student attitude variables.

4.7. DISCUSSION

The results presented in this chapter make an important contribution to our understanding of the sources of South Africa's pro-girl advantage in reading and mathematics achievement in the primary school grades. By employing a decomposition approach to the country's PIRLS and TIMSS results, I am able to separate gender gaps in reading and mathematics achievement into their explained and unexplained components, which provides some indication of the observable differences between boys and girls, and differences in how those characteristics are translated into achievement, that may contribute to achievement gaps in learning outcomes.

4.7.1. SUMMARY OF MAIN RESULTS

Three main results emerge from the analysis presented here. Firstly, in terms of the main research question, the results from the first set of decompositions indicate that around 40% of the pro-girl gap in PIRLS and around 50% of the pro-girl gap in TIMSS can be explained by differences in endowments of observable characteristics between boys and girls captured in these datasets. The magnitudes of these explained components of the gender gap in achievement are similar to those found in studies conducted in other countries that employ a similar methodology (see for example Badr, Morrissey and Appleton (2012); Cobb-Clark and Moschion (2017)). When accounting for gendered repetition patterns in prior grades, however, the size of the explained component of the gender gap decreases significantly in both PIRLS and TIMSS. This effect is particularly pronounced in TIMSS, where the explained component of the gender gap is reduced to insignificance in two of the four quintiles with a significant pro-girl achievement gap in the second set of decompositions. This result provides evidence in support of the hypothesis posited in the introduction, namely that part of the observed pro-girl achievement gaps in these datasets may be due to a selection effect whereby a given Grade 4 or 5 class consists of a larger proportion of boys than girls who have repeated a grade. This, in turn, suggests that the pro-girl achievement gaps observed in PIRLS and TIMSS are partly the result of a pro-girl advantage that is already evident at the start of formal schooling. Importantly, however, the second set of decompositions show that large pro-girl achievement gaps remain in both PIRLS and TIMSS, even when limiting the samples to only students who are on-track in terms of age-for-grade. In other words, while the first set of decompositions show that girls' advantage in terms of prior grade completion does contribute to the observed pro-girl gap in both PIRLS and TIMSS, the results from the second set of decompositions of the performance of on-track students indicate that girls' advantage in the prior grades only *partly* explains the pro-girl gaps in these datasets.

Secondly, the results from the second set of decompositions (for on-track students) suggest that while differences in boys' and girls' attitudes toward reading and school may contribute to the observed pro-girl achievement gap in PIRLS, even when considering only students who are on-track in terms of age, this is not the case for TIMSS. This constitutes a noteworthy result since it suggests there is an element of domain specificity to the role of student attitudes in South Africa's pro-girl advantage in educational outcomes, whereby gender differences in the same set of subject-specific attitudes do not contribute equally to girls' advantage in TIMSS mathematics achievement as they do to girls' advantage in PIRLS reading achievement. Specifically, girls' higher reading self-concept as well as their higher enjoyment of reading contribute significantly to the pro-girl gap in PIRLS across virtually all school quintiles. By contrast, even though girls in TIMSS have more positive attitudes toward mathematics than boys, these differences do not significantly contribute to the pro-girl advantage in TIMSS scores. This is in line with existing evidence in the international literature that the nature of gendered educational outcomes differs across domains, especially reading and mathematics (Cobb-Clark and Moschion, 2017).

Lastly, I present evidence that the magnitude of the pro-girl gaps in both PIRLS and TIMSS are related to the SES of the school, with larger gaps observed in the lower school quintiles. This effect is particularly pronounced in TIMSS, where splitting the sample into school SES quintiles reveals that the pro-girl gap observed for the full sample masks the fact that there is no significant pro-girl gap in the wealthiest 20% of schools. Similarly, I find that there are SES differences in the significance of student attitudes in explaining the pro-girl gaps in PIRLS. Together, these results suggest that gender, SES, and student attitudes all interact to produce learning outcomes, something that has received scant attention in the local literature to date. This result suggests that interrogating the intersection between gender, SES and student attitudes further is likely to add more nuance to our understanding of how these factors interact to produce learning outcomes in South Africa more generally.

4.7.2. LIMITATIONS

While these results make an important contribution to our understanding of South Africa's pro-girl advantage in educational outcomes, the analysis presented here is subject to a number of limitations. Firstly, the results presented here are subject to the same limitations that plague all studies that employ Oaxaca-Blinder decomposition analysis (see Fortin, Lemieux and Firpo (2011) for a comprehensive discussion of these issues). Perhaps most pressingly, interpreting the unexplained component of the gender gap as a response effect requires the assumption that all factors that matter for achievement are included in the decomposition model (Jann, 2008). This is of course a very difficult condition to meet, and consequently the nature of the unexplained components of the gender gaps was deliberately de-emphasised in this chapter. While the results presented here of the role of endowment differences between boys and girls in explaining South Africa's pro-girl gap in PIRLS and TIMSS make an important contribution to the literature, there is still much that we do not know regarding potential differences in how the endowments of boys and girls are translated into achievement. Given evidence in the international literature of such effects, future research in South Africa would do well to investigate this further.

A second major limitation lies in the construct validity of the measures aimed at tapping student attitudes toward school and learning. Importantly, the psychometric properties of these measures have not been established in the South African context, and existing evidence from PIRLS and TIMSS data from other countries suggests these measures may be subject to a number of limitations, including method and translation effects (see for example Marsh *et al.* (2013); Bofah and Hannula (2015); and Alghamdi (2018)). Moreover, there may be important differences in how boys and girls respond to the questionnaire items aimed at identifying these constructs. Despite these limitations, I maintain that these measures do capture something about the more affective aspects of learning that meaningfully contributes to South Africa's pro-girl achievement gap. Another important avenue for future research would be further investigating the role of student attitudes, and non-cognitive skills more generally, in contributing to South Africa's pro-girl gap in education.

Thirdly, the PIRLS and TIMSS assessments are by grade and not by age (such as, for example, the Programme for International Student Assessment (PISA), which tests all 15-year-olds regardless of what grade they are in), thus it does not allow for analysis of all members of the same cohort. Limiting the sample to students who are on-track in terms of age-for-grade allows one to get closer to the 'true' cohort that started school together in Grade 1. However the sample of students remains a pseudo cohort, since it includes students from a previous cohort (students who started school early and repeated a grade), or excludes students from the true cohort (students who started school late and never repeated). This limitation points to the need for international educational assessments to include items in the student background questionnaire that would provide a more detailed picture of students' past educational trajectories, such as asking students whether they have ever repeated a grade.

Lastly, while the results of this study highlight the importance of gendered repetition patterns in the early grades in contributing to South Africa's observed pro-girl achievement gap in Grade 4 reading and Grade 5 mathematics, this result only shifts the question of the sources of the pro-girl advantage to the early grades. That is, the results presented here do not bring us any closer to understanding *why* boys are underachieving relative to girls in the early grades. To my knowledge, there is only one published study that investigates gender differences in early learning outcomes in South Africa: that of Wilsenach and Makaure (2018), who investigate gender differences in phonological processing skills among Northern Sotho-speaking Grade 3 students, and present evidence of a pro-girl advantage in these skills. The results from the decomposition analysis in this study suggest much more needs to be done to understand the reasons behind boys' disadvantage in the early grades, since much of the pro-girl achievement gap in Grade 4 reading and Grade 5 mathematics can be attributed the pro-girl advantage in grade completion in the early grades.

4.8. CONCLUSION

This study contributes to the literature documenting South Africa's pro-girl advantage in educational outcomes, and adds to our understanding of the sources of the pro-girl achievement gaps in Grade 4 reading and Grade 5 mathematics by decomposing these gaps into their explained and unexplained components, respectively. The analysis in this chapter focussed on three aspects of the pro-girl achievement gap that has hitherto received little attention, namely the potential role of gendered repetition patterns in the foundation phase in contributing to the observed pro-girl gap in PIRLS and TIMSS; differences in the magnitude and potential sources of this gap among students from different socioeconomic backgrounds; and the contribution of gender differences in student attitudes to this gap. The results of the analysis suggest that different processes are at play in the production of South Africa's pro-girl gap in reading and mathematics achievement, and for students from different socioeconomic backgrounds. Although the results presented here constitute an important first step towards understanding the sources of South Africa's pro-female advantage in educational outcomes, we need to do more to understand why this gap is so pronounced and persistent across the education system, with research that focusses on gender gaps in the early grades likely constituting a particularly fruitful avenue for future research.

CHAPTER 5: CONCLUSION

The research presented in this thesis has investigated the role of non-cognitive skills in educational production in South Africa. This question was motivated by the need to add evidence from a LMIC to the international evidence base of the association between non-cognitive skills and student achievement. Non-cognitive skills have become central to education research and policy debates in HICs, but an important unanswered question in the international literature is whether the strong association between non-cognitive skills and educational outcomes observed in HICs holds LMIC contexts, which are often characterised by severe resource deprivation. The three chapters in this thesis aimed to address this gap by investigating the relationship between non-cognitive skills and learning outcomes in South Africa, with a view to both improve our understanding of the educational production process in South Africa, as well as to add to the international evidence base regarding the role of non-cognitive skills in educational production in LMICs more generally.

5.1. SUMMARY OF MAIN FINDINGS

This thesis investigated the potential role of non-cognitive skills in producing learning outcomes in South Africa. Chapters 2, 3, and 4 each considered this question through a different lens, which informed the econometric strategy employed. Four datasets were utilised to model the production of student achievement in two subjects (reading and mathematics) at different grade levels. This allowed for an investigation into the relationship between non-cognitive skills and learning outcomes across different subjects and in different grades.

Chapter 2, "Performance Beyond Expectations: Examining correlates of exceptional academic achievement in high-poverty contexts in South Africa", made use of South African PIRLS and TIMSS data to investigate what makes some students achieve Grade 4 reading and Grade 9 mathematics results beyond expectations. It is argued that the South African context is particularly suited for investigating this question, given that many children in South Africa face severely under-resourced home and school environments that are akin to those faced by children in much of the developing world. The first research objective in this chapter was to identify exceptional performers in the PIRLS 2016 and TIMSS 2015 datasets. This study considered how these students are distributed across schools of differing quality, and how they perform relative to the median student in their school. This was done in order to determine whether the phenomenon whereby some students achieve good academic results despite contexts of socio-economic deprivation occurs at the level of the individual child or the school. That is, I endeavoured to investigate whether the outlier students observed in the PIRLS and TIMSS data are concentrated in a small number of schools (an 'outlier school' hypothesis), or whether these students are scattered across schools of differing quality (an 'outlier child' hypothesis). The second research objective was to determine the ways in which these students differ systematically from their lower-achieving peers. This was achieved using logistic ordinary least squares (OLS) regression analysis to investigate factors at the individual, family, and school level that contribute to the probability that a student will achieve exceptional academic results.

The results from Chapter 2 indicated that exceptional performers are scattered across schools of varying quality, suggesting that the phenomenon whereby some students achieve good academic results despite contexts of socio-economic deprivation occurs at the level of the individual child, and not the school. That is, students from poor home backgrounds who achieve good academic results are outliers in their schools, constituting evidence in support of an ‘outlier child’ hypothesis, as opposed to an ‘outlier school’ hypothesis. This result echoes findings from Wills (2017) and Spaull and Pretorius (2019), and strongly suggests that there are important individual and home background characteristics that enable some students to achieve good PIRLS reading and TIMSS mathematics results, despite attending schools with very low average levels of performance. This is a noteworthy result since it points to an important feature of the educational production process, highlighted by Hanushek (1979), that schools do not have homogenous effects on students, and that individual characteristics of students are important inputs into the process of educational production. In particular, I find that the constructs in the PIRLS and TIMSS data aimed at capturing student attitudes toward reading and mathematics are strong predictors of exceptional performance in these assessments, respectively. This finding adds to the growing international evidence base of a strong association between non-cognitive skills and academic achievement. Moreover, the finding that student attitudes are highly predictive of exceptional academic performance in the South African context comprises a particularly noteworthy contribution to the international literature on the role of non-cognitive skills in educational production, which is dominated by evidence from HICs.

Another noteworthy result from the multivariate analysis in Chapter 2, which emerges again in both Chapters 3 and 4, is that the strength of the association between non-cognitive skills and academic achievement differs by the socio-economic context of the school. This makes an important contribution to the literature on the role of non-cognitive skills in educational production, which has seldom considered the interaction of contextual factors and non-cognitive skills in producing learning outcomes. More broadly, this result contributes to the international economics of education literature by providing evidence of contextual differences in the association between educational inputs (in this case, non-cognitive skills) and student achievement, and supports the emergent perspective within our discipline that contextual factors interact with given educational inputs to produce learning outcomes (Pritchett and Sandefur, 2013; Ravallion, 2020; Hanushek, 2021).

Building on these results, Chapter 3, “Perseverance, Passion, and Poverty: Examining the association between grit and reading achievement in high-poverty schools in South Africa” explicitly investigates whether school functionality moderates the association between the non-cognitive skill of ‘grit’ and

Grade 6 reading achievement in schools characterized by contexts of severe resource deprivation. Data from the Leadership for Literacy study, containing information about more than 2,600 students from 60 township and rural schools across three provinces in South Africa, was used to answer this question. While Chapter 2 focused on exceptional performers, Chapter 3 explored the relationship between grit and achievement across the spectrum of reading performance in the Leadership for Literacy data. The results from this chapter add to the evidence presented in Chapter 2 of a strong association between non-cognitive skills and academic performance, even in contexts characterized by severe resource deprivation, such as township and rural areas in South Africa. Notably, I find that the perseverance subscale of grit is the strongest predictor of reading comprehension test scores out of all the covariates included in the OLS regression – including a host of factors at the home and school level. This is a noteworthy result, since it suggests that non-cognitive skills do not only have a strong association with student performance in high-poverty contexts, but perseverance is *the* strongest observed predictor of performance out of all the factors included in the model.

The econometric analysis in Chapter 3 further points to potential moderating effects between the perseverance subscale of grit and school functionality in predicting student achievement. This result makes an important contribution to the international literature on non-cognitive skills in education, since it is one of only a handful of studies that investigates whether other educational inputs interact with non-cognitive skills to predict learning outcomes, and the first to investigate whether school quality, specifically, interacts with the non-cognitive skill of grit. The results from Chapter 3 also suggest that that the nature of the interaction between perseverance and school functionality is not uniform across the distributions of these variables. In terms of the broader literature on potential interaction effects between non-cognitive skills and other educational inputs, this is an important result, since it provides empirical evidence for the theoretical possibility that the nature of interaction effects between non-cognitive skills and other educational inputs may vary at different points of the distributions of both non-cognitive skills and the inputs they interact with.

Given the links that have been made between non-cognitive skills and gendered education outcomes in the international literature, a natural extension of the results from Chapters 2 and 3 was exploring whether South Africa's gendered educational outcomes can be linked to gender differences in non-cognitive skills. This question is investigated in Chapter 4, "South Africa's pro-girl advantage in PIRLS and TIMSS: How much can be explained?" It is argued that the role of non-cognitive skills in contributing to gendered education outcomes deserves particular attention in South Africa, given that the country exhibits one of the largest pro-girl advantages in education in the world. To investigate this, the PIRLS and TIMSS data was used to model student achievement, this time using the Grade 5 TIMSS data to model mathematics achievement. Oaxaca-Blinder decomposition analysis was used to investigate how much of South Africa's pro-girl achievement gap in these datasets can be explained by gender differences in observable characteristics, with a particular focus on the contribution of gender differences in non-cognitive skills.

The contributions of Chapter 4 are twofold. Firstly, analysis of the age distribution of students in both the Grade 4 PIRLS and Grade 5 TIMSS data indicates that boys are much more likely to be older than the appropriate age for their grade, suggesting that boys are more likely to repeat a year in the foundation phase (Grades 1-3). This, in turn, suggests that part of South Africa's pro-girl advantage in Grade 4 reading and Grade 5 mathematics achievement can be explained by differential grade repetition patterns in the foundation phase. This result points to the conclusion that the observed pro-girl advantage in PIRLS and TIMSS finds its roots in the foundation phase – or even before that.

The results presented in Chapter 4 further suggest, however, that gendered repetition patterns in the foundation phase are only part of the story, in terms of explaining South Africa's pro-girl advantage in PIRLS and TIMSS. A large pro-girl gap remains in both datasets, even after accounting for these gendered patterns in the proportion of students who are over-age for their grade. In accordance with evidence from the international literature, the results suggest that a significant proportion of the pro-girl achievement gap in Grade 4 reading can be attributed to the fact that girls have more positive attitudes toward reading than boys. This result adds to the evidence presented in Chapters 2 and 3 of the strong association between non-cognitive skills and academic achievement. Furthermore, this result builds on this evidence by showing that gender differences in non-cognitive skills partly explain South Africa's pro-girl advantage in Grade 4 reading achievement. An interesting, and perhaps unexpected, result from Chapter 4 is that gender differences in attitudes toward mathematics do not significantly contribute to the observed pro-girl achievement gap in mathematics in TIMSS, constituting evidence that the association between non-cognitive skills and the pro-girl achievement gap is not uniform across subjects. Overall, the analysis in this chapter illustrates clearly how focusing on non-cognitive skills as predictors of learning outcomes can enhance our understanding of hitherto unexplained features of South Africa's educational performance, such as the country's large and persistent pro-girl achievement gap.

5.2. LIMITATIONS

The results across Chapters 2, 3, and 4 are subject to a number of important limitations which can be grouped into three main areas of concern, namely (i) lack of conceptual clarity around what non-cognitive skills represent, (ii) measurement challenges, and (iii) endogeneity issues.

5.2.1. LACK OF CONCEPTUAL CLARITY AROUND WHAT NON-COGNITIVE SKILLS REPRESENT

A major limitation of studies which use large-scale data to investigate the relationship between non-cognitive skills and learning outcomes (including the three studies in this thesis) is that the choice of non-cognitive skills is often based on data availability, rather than solid theoretical underpinnings (Farrington *et al.*, 2012; Garcia, 2013; Scorza *et al.*, 2017). The measures of student self-confidence, engagement, and belonging in the PIRLS and TIMSS datasets are widely used in the literature; however there is no evidence that these measures are either most pertinent to school performance or most malleable in school settings. In a meta-analysis of the grit literature, Credé (2018: 5) argues that this literature "is currently characterised by a number of serious theoretical and empirical challenges", and highlights the overlap between grit and the existing construct of conscientiousness, as well as the lack of evidence regarding the malleability of grit through targeted interventions, as particularly problematic.

These criticisms highlight the lack of conceptual clarity around many of the constructs that are included under the broad umbrella of 'non-cognitive skills' in education research. This limitation is hardly surprising, given the generic definition of non-cognitive skills that is currently used in the literature. According to Heckman and Kautz (2012: 2), for example, non-cognitive skills include "personality traits, goals, motivations, and preferences". In this sense, the term non-cognitive skills is used as a catch-all term for everything that matters for achievement and is not captured by measures of 'cognitive' skills, such as IQ. Such a conceptualisation limits the applicability of studies of non-cognitive skills to education policy and practice. This limitation will have to be addressed in future research if we are to advance the science of non-cognitive skills in education.

5.2.2. MEASUREMENT CHALLENGES

A related limitation, raised throughout this thesis, is the reliability with which non-cognitive skills are measured. The concerns around self-reported measures of non-cognitive skills are well-documented in the literature from industrialised countries (Marsh *et al.*, 2013; Duckworth and Yeager, 2015; Halle and Darling-Churchill, 2016; Del Bono, Kinsler and Pavan, 2019). These concerns are even more pressing in contexts of low literacy, such as South Africa, and many other developing countries. While attempts were made to reduce bias resulting from measurement error as far as possible throughout this thesis, the lack of reliable measures of non-cognitive skills, especially in low-literacy settings, remains a major concern. Another important step in advancing the study of non-cognitive skills in education research will, therefore, be developing better measures of non-cognitive skills.

5.2.3. ENDOGENEITY

The non-experimental nature of the data used in this thesis inevitably means that the evidence presented of a positive association between non-cognitive skills and achievement rests on the strong assumption that all factors that matter for student achievement are included in the models of educational production. While this is a major limitation, it is one that plagues most studies in the production function tradition, barring very recent contributions that propose complex econometric techniques for eluding this problem (see for example Agostinelli, Saharkhiz and Wiswall, 2019). Crucially, these techniques have strong data requirements that challenge the feasibility of applying them to data from countries such as South

Africa, including detailed longitudinal measures of students' cognitive and non-cognitive skills. Therefore, it is maintained that while the results presented in this thesis are subject to the host of endogeneity problems that are associated with making inferences from observational data, the results presented nonetheless make a noteworthy contribution to both the local and international literature by adding evidence from South Africa to the international evidence base of the association between non-cognitive skills and student achievement.

5.3. RECOMMENDATIONS

The analyses and results in this thesis suggest that the role of non-cognitive skills in educational production in LMICs such as South Africa is a promising area for future research. This research should focus on addressing the limitations set out in the section above, especially if we intend to take seriously the possibility that improving students' non-cognitive skills may be a powerful but hitherto overlooked policy lever for raising learning outcomes. This subsection summarises three recommendations for future research on the role of non-cognitive skills in educational production. This is done with a view to mapping a research agenda for the study of non-cognitive skills in the production of learning outcomes internationally, with the ultimate goal of informing policy aimed at improving learning outcomes, particularly in LMICs.

5.3.1. EVIDENCE FROM MORE LMICS

It is clear from the evidence presented in this thesis that studying non-cognitive skills in an LMIC like South Africa allows us to investigate important questions about the role of these skills in education that evidence from HICs is not able to address. In particular, the question of well non-cognitive skills predict student achievement in contexts of severe resource deprivation and poor school functionality can only be investigated in countries whose education systems are characterised as such. Adding evidence from LMICs to the international evidence base may therefore help us improve existing theories of the role of non-cognitive skills in education. Studying these skills in LMICs therefore has the potential to benefit not only research and practice in these countries, but our discipline as a whole. There is ample evidence that, at least for the time being, our discipline is committed to better understanding the role of non-cognitive skills in education. The past decade of economics research has seen not only a surge in books and journal articles on non-cognitive skills, but also the establishment of working groups and entire research units at some of the top institutions in our field devoted to this topic. The establishment of the Centre for the Economics of Human Development at the University of Chicago in 2014 is a prime example of the centrality of the study of non-cognitive skills to our discipline[34]. The establishment of the Human Capital and Economic Opportunity (HCEO) Global Working Group in 2010 is another prime example. The working group consists of more than 500 researchers (including some of the top

[34] See the CEHD's website: https://cehd.uchicago.edu.

researchers in our discipline, such as Hanushek and Heckman) and focusses on promoting interdisciplinary research with a view to improving our understanding of human capital development (including non-cognitive skills) and its impact on inequality (HCEO, n.d.). It is thus clear that our discipline is devoting significant funding and research effort to understanding the role of non-cognitive skills in predicting meaningful life outcomes. The evidence presented in this thesis suggests studying non-cognitive skills in an LMIC like South Africa may provide new insights that can advance this research agenda, even with relatively simple econometric techniques and limited data. As such, further studying non-cognitive skills in LMICs could be "low-hanging fruit" in the broader research project that our discipline is already devoting much funding and research effort to.

5.3.2. ALIGNING FRAMEWORKS ACROSS DISICPLINES

Osher *et al.* (2017) argue that gaining conceptual clarity around what non-cognitive skills represent will require, first and foremost, aligning the various different frameworks that are currently used to study the role of non-cognitive skills in education. They argue that the focus of this alignment should be creating a common and clear language of non-cognitive skills for practitioners. An example of efforts toward such alignment is the Collaborative for Academic, Social, and Emotional Learning's (CASEL) establishment of the Establishing Practical Social-Emotional Competence Assessments Work Group. The work group consists of over 60 researchers and practitioners and has released a number of synthesis reports that aim to create the alignment in research on non-cognitive skills in education that Osher *et al.* (2017) call for (see for example Berg *et al.* (2019) and McKown (2019)). Unfortunately, CASEL's work group focuses exclusively on the development of aligned frameworks for the study of non-cognitive skills for the North American context. The establishment of a similar working group for research on non-cognitive skills in LMICs specifically is likely to promote alignment in this research at the outset.

5.3.3. IMPROVING MEASUREMENT

In addition to efforts aimed at better defining non-cognitive skills, the limitations related to the measurement of non-cognitive skills raised throughout this thesis points to the importance of devoting significant research efforts to developing better measures of these skills. Of particular importance is the need to develop instruments that can reliably measure non-cognitive skills in LMIC contexts. This is a small but growing area of research within the international development literature (see for example the work of Laajaj and Macours (2017) and Jukes *et al.* (2018)) that deserves more attention if we are to better understand the role of non-cognitive skills in educational production in LMICs, with the view to designing policy that can effectively raise learning outcomes in these countries.

5.4. CONCLUDING REMARKS

This thesis has intended to improve our understanding of how learning outcomes are produced in South Africa by investigating the role of non-cognitive skills in this process. This research question was motivated by the centrality of non-cognitive skills to education research and policy internationally, and the dearth of evidence from LMICs regarding the role of non-cognitive skills in producing learning outcomes. In accordance with evidence from HICs, the results from all three chapters in this thesis point to a strong association between non-cognitive skills and academic achievement in South Africa. This evidence makes a noteworthy contribution to the international literature, since it suggests non-cognitive skills are an important input in the production of learning outcomes, even in contexts of severe resource deprivation. This suggests that further investigating the role of non-cognitive skills in producing learning outcomes in LMICs such as South Africa is a promising area for future research. In addition, the analyses in this thesis consider whether socio-economic and school contexts moderate the association between non-cognitive skills and student achievement – that is, whether this association differs across socio-economic and school contexts. The potential interaction of contextual factors and other educational inputs (in this case, non-cognitive skills) is an under-researched yet important question for international education research and practice (Pritchett and Sandefur, 2013; Ravallion, 2020; Hanushek, 2021b). The fact that the results presented in this thesis provide evidence of contextual differences in the association between non-cognitive skills and student achievement even within the same country suggests that further exploring such interaction effects is also likely to be a fruitful area for future research.

REFERENCES

Aberra, R. A. (2018) 'Cognitive and non-cognitive gender gap among children: Evidence from Ethiopia', *Socialni Studia/Social Studies*, 15(1), pp. 101–119. doi: 10.5817/soc2018-1-101.

Abu-Hilal, M. M. *et al.* (2013) 'Construct validity of self-concept in TIMSS's student background questionnaire: A test of separation and conflation of cognitive and affective dimensions of self-concept among Saudi eighth graders', *European Journal of Psychology of Education*, 28(4), pp. 1201–1220. doi: 10.1007/s10212-012-0162-1.

Agasisti, T. *et al.* (2018) *Academic Resilience: What schools and countries do to help disadvantaged students succeed in PISA*. OECD Discussion Paper No. 167. Paris.

Agasisti, T. and Longobardi, S. (2017) 'Equality of Educational Opportunities, Schools' Characteristics and Resilient Students: An Empirical Study of EU-15 Countries Using OECD-PISA 2009 Data', *Social Indicators Research*. Springer Netherlands, 134(3), pp. 917–953. doi: 10.1007/s11205-016-1464-5.

Agostinelli, F. (2018) *Essays on Children's Skill Formation*. Doctoral dissertation. Arizona State University. doi: 10.1017/CBO9781107415324.004.

Agostinelli, F., Saharkhiz, M. and Wiswall, M. J. (2019) Home and School in the Development of Children. *NBER Working Papers*, No. 26037. Cambridge, MA.

Alghamdi, M. H. (2018) *Multilevel Student Engagement and Confidence in Learning: Assessing Construct Validity Across Students and Schools*. Master's thesis. University of Kansas.

Almlund, M. *et al.* (2011) 'Personality Psychology and Economics', in *Handbook of the Economics of Education*. Cambridge, pp. 1–181. doi: 10.1016/B978-0-444-53444-6.00001-8.

Anderson, K. G., Case, A. and Lam, D. (2001) 'Causes and Consequences of Schooling Outcomes in South Africa', *Social Dynamics*, 27(January 2015), pp. 37–59. doi: 10.1080/02533950108458703.

Andrabi, T., Daniels, B. and Das, J. (2020) *Human Capital Accumulation and Disasters: Evidence from the Pakistan Earthquake of 2005*. 20/039. doi: 10.35489/BSG-RISE.

Andreoni, J. *et al.* (2019) 'Risk Preferences of Children and Adolescents in Relation to Gender, Cognitive Skills, Soft Skills, and Executive Functions', *Journal of Economic Behavior and Organization*, 143, pp. 28–44.

Attanasio, O. *et al.* (2020) 'Estimating the production function for human capital: Results from a randomized controlled trial in Colombia', *American Economic Review*, 110(1), pp. 48–85. doi: 10.1257/aer.20150183.

Autor, D. *et al.* (2016) 'School quality and the gender gap in educational achievement', *American Economic Review*, 106(5), pp. 289–295. doi: 10.1257/aer.p20161074.

Azina, I. N. and Halimah, A. (2012) 'Student factors and mathematics achievement: Evidence from TIMSS 2007', *Eurasia Journal of Mathematics, Science and Technology Education*, 8(4), pp. 249–255. doi: 10.12973/eurasia.2012.843a.

Badr, M., Morrissey, O. and Appleton, S. (2012) 'Gender differentials in maths test scores in MENA countries', *CREDIT Research Paper*.

Beattie, G., Laliberté, J. P. and Oreopoulos, P. (2016) *Thrivers and Divers: Using Non-Academic Measures to Predict College Success and Failure*, *NBER Working Papers*. 22629. doi: 10.1017/CBO9781107415324.004.

Berg, J. *et al.* (2019) *Socio-Emotional Competencies in Context: Using Social-Emotional Learning Frameworks to Build Educators' Understanding*, *Framework Briefs, Comparative Series*. 2. doi: 10.1080/10888691.2017.1398650.

Bertrand, M. and Pan, J. (2011) The Trouble With Boys: Social Influences and the Gender Gap in Disruptive Behaviour. *NBER Working Paper*, No. 17541. Cambridge, MA. doi: 10.16194/j.cnki.31-1059/g4.2011.07.016.

Bofah, E. A. tutu and Hannula, M. S. (2015) 'TIMSS data in an African comparative perspective: Investigating the factors influencing achievement in mathematics and their psychometric properties', *Large-Scale Assessments in Education*. Springer US, 3(1). doi: 10.1186/s40536-015-0014-y.

Borman, G. D. and Overman, L. T. (2004) 'Academic Resilience in Mathematics among Poor and Minority Students', *The Elementary School Journal*, 104(3), pp. 177–195.

Borman, G. and Rachuba, L. (2001) *Academic success among poor and minority students: An analysis of competing models of school effects*, *Center for Research on the Education of Students Placed At Risk*. Available at: https://eric.ed.gov/?id=ED451281.

Bowles, S. and Gintis, H. (1976) *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. 1st edn. Edited by S. Bowles and H. Gintis. Chicago: Haymarket Books.

Branson, N., Hofmeyr, C. and Lam, D. (2014) 'Progress through school and the determinants of school dropout in South Africa Progress through school and the determinants of school dropout in South Africa', *Development Southern Africa*, 31(September), pp. 106–126. doi: 10.1080/0376835X.2013.853610.

Branson, N. and Lam, D. (2009) Education: Analysis of the NIDS Wave 1 Dataset, Social Development. 3. *SALDRU Working Papers*. Cape Town.

Branson, N. and Lam, D. (2010) 'Education inequality in south africa: evidence from the national income dynamics study', *Studies in Economics and Econometrics*, 34(3), pp. 85–109. Available at: http://reference.sabinet.co.za/document/EJC21505.

Burger, R. (2011) 'School effectiveness in Zambia: The origins of differences between rural and urban outcomes', *Development Southern Africa*, 28(2), pp. 157–176. doi: 10.1080/0376835X.2011.570064.

Cappella, E. and Weinstein, R. S. (2001) 'Turning around reading achievement: Predictors of high school students' academic resilience', *Journal of Educational Psychology*, 93(4), pp. 758–771. doi: 10.1037/0022-0663.93.4.758.

Carey, L. M. (2001) *Measuring and Evaluating School Learning*. 3rd edn. Boston: Allyn and Bacon.

Case, B. A. and Deaton, A. (2005) 'Health and Wealth among the Poor: India and South Africa Compared', *American Economic Review*, 95(2), pp. 229–233.

Cheung, K. (2016) 'The effects of resilience in learning variables on mathematical literacy performance: a study of learning characteristics of the academic resilient and advantaged low achievers in Shanghai, Singapore, Hong Kong, Taiwan and Korea', *Educational Psychology*, 37(8), pp. 965–982. Available at: https://www-tandfonline-com.ez.sun.ac.za/doi/abs/10.1080/01443410.2016.1194372.

Cobb-Clark, D. A. and Moschion, J. (2017) 'Gender gaps in early educational achievement', *Journal of Population Economics*. Journal of Population Economics, 30(4), pp. 1093–1134. doi: 10.1007/s00148-017-0638-z.

Coleman, J. S. and Others, A. (1966) 'Equality of Educational Opportunity.' Available at: http://eric.ed.gov/?id=ED012275.

Connell, J. P., Spencer, M. B. and Aber, J. L. (1994) 'Educational Risk and Resilience in African-American Youth: Context, Self, Action, and Outcomes in School', *Child Development*, 65, pp. 493–506.

Cordero, J. M., Cristóbal, V. and Santín, D. (2018) 'Causal Inference on Education Policies: a Survey of Empirical Studies Using Pisa, Timss and Pirls', *Journal of Economic Surveys*, 32(3), pp. 878–915. doi: 10.1111/joes.12217.

Cosgrove, J. and Creaven, A.-M. (2013) 'Understanding achievement in PIRLS and TIMSS 2011', in Eivers, E. and Clerkin, A. (eds) *National Schools, International contexts: Beyond the PIRLS and TIMSS test results*. Dublin: Educational Research Centre, pp. 201–240.

Credé, M. (2018) 'What Shall We Do About Grit? A Critical Review of What We Know and What We Don't Know', *Educational Researcher*, 47(9), pp. 606–611. doi: 10.3102/0013189X18801322.

Credé, M., Tynan, M. C. and Harms, P. D. (2017) 'Much ado about grit: A meta-analytic synthesis of the grit literature', *Journal of Personality and Social Psychology*, 113(3). doi: 10.1037/pspp0000102.

Cunha, F. and Heckman, J. (2007) 'The technology of skill formation', *American Economic Review*, 97(2), pp. 31–47. doi: 10.1257/aer.97.2.31.

Cunha, F. and Heckman, J. J. (2008) 'Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation', *Journal of Human Resources*, 43(4), pp. 738–782. doi: 10.3368/jhr.43.4.738.

Damian, R. I. *et al.* (2015) 'Can Personality Traits and Intelligence Compensate for Background Disadvantage? Predicting Status Attainment in Adulthood', *Journal of Personality and Social Psychology*, 109(3), pp. 473–489. doi: 10.1016/j.physbeh.2017.03.040.

Das, D. (2018) 'Academic Resilience Among Children from Disadvantaged Social Groups in India', *Social Indicators Research*. Springer Netherlands. doi: 10.1007/s11205-018-1899-y.

Del Bono, E. Del, Kinsler, J. and Pavan, R. (2019) 'Skill Formation and the Trouble with Child Non-Cognitive Skill Measures'. *IZA Discussion Papers*. Bonn, Germany.

Delprato, M., Akyeampong, K. and Dunne, M. (2017) 'The impact of bullying on students' learning in Latin America: A matching approach for 15 countries', *International Journal of Educational Development*. Elsevier Ltd, 52, pp. 37–57. doi: 10.1016/j.ijedudev.2016.10.002.

Dercon, S. and Singh, A. (2013) 'From Nutrition to Aspirations and Self-Efficacy: Gender Bias over Time among Children in Four Countries', *World Development*. Elsevier Ltd, 45, pp. 31–50. doi: 10.1016/j.worlddev.2012.12.001.

Diaz, J. J., Arias, O. and Tudela, D. V. (2013) *Does Perseverance Pay as Much as Being Smart?: The Returns to Cognitive and Non-cognitive Skills in urban Peru*. Washington, D.C.

DiPrete, T. A. and Jennings, J. L. (2012) 'Social and behavioral skills and the gender gap in early educational achievement', *Social Science Research*, 41(1), pp. 1–15. doi: 10.1016/j.ssresearch.2011.09.001.

Duckworth, A. L. *et al.* (2007) 'Grit: Perseverance and Passion for Long-Term Goals', *Journal of Personality and Social Psychology*, 92(6), pp. 1087–1101. doi: 10.1037/0022-3514.92.6.1087.

Duckworth, A. L. and Quinn, P. D. (2009) 'Development and validation of the short Grit Scale (Grit-S)', *Journal of Personality Assessment*, 91(2), pp. 166–174. doi: 10.1080/00223890802634290.

Duckworth, A. L. and Seligman, M. E. P. (2006) 'Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores', *Journal of Educational Psychology*, 98(1), pp. 198–208. doi: 10.1037/0022-0663.98.1.198.

Duckworth, A. L. and Yeager, D. S. (2015) 'Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes', *Educational Researcher*, 4(May), pp. 237–251. doi: 10.3102/0013189X15584327.

Duncan, G. J. and Magnuson, K. (2011) 'The Nature and Impact of Early Achievement Skills, Attention Skills, and Behaviour Problems', in Duncan, G. J. and Murnane, R. J. (eds) *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. 1st edn. New York: Russell Sage Foundation, pp. 47–69.

Durlak, J. A. J. A. *et al.* (2011) 'The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions', *Child Development*, 82(1), pp. 405–432. doi: 10.1111/j.1467-8624.2010.01564.x.

Dweck, C., Walton, G. M. and Cohen, G. L. (2011) *Academic tenacity: Mindset and skills that promote long-term learning*. Seattle. Available at: http://web.stanford.edu/~gwalton/home/Welcome_files/DweckWaltonCohen_2014.pdf.

Egalite, A. J., Mills, J. N. and Greene, J. P. (2016) 'The softer side of learning: Measuring students' non-cognitive skills', *Improving Schools*, 19(1). doi: 10.1177/1365480215616313.

Engzell, P. (2019) 'What Do Books in the Home Proxy For? A Cautionary Tale', *Sociological Methods and Research*, pp. 1–28. doi: 10.1177/0049124119826143.

Entwisle, D. R., Alexander, K. L. and Olson, L. S. (2007) 'Early schooling: The handicap of being poor and male', *Sociology of Education*, 80(2), pp. 114–138. doi: 10.1177/003804070708000202.

Erberber, E. *et al.* (2015) 'Socioeconomically disadvantaged students who are academically successful: Examinig academic resilience cross-nationally', *IEA's Policy Brief Series*.

Espinosa, A. M. G. (2017) *Estimating the Education Production Function for Cognitive and Non-cognitive development of children in Vietnam through structural equation modelling using Young Lives Data Base*. London.

Evans, D. K. and Yuan, F. (2019) *Equivalent Years of Schooling A Metric to Communicate Learning Gains in Concrete Terms*, *World Bank Policy Research Working Paper*. Available at: www.worldbank.org/research.

Farkas, G. (2011) 'Middle and High School Skills, Behaviours, Attitudes, and Curriculum Enrolment, and Their Consequences', in Duncan, G. J. and Murnane, R. J. (eds) *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*. New York: Russell Sage Foundation, pp. 72–90.

Farrington, C. A. *et al.* (2012) 'Teaching Adolescents To Become Learners: The Role of Noncognitive Factors in Shaping School Performance', *University of Chicago Consortium on Chicago School Research*, (June), pp. 1–106.

Finn, J. D. and Rock, D. A. (1997) 'Academic Success Among Students at Risk for School Failure', *Journal of Applied Psychology*, 82(2), pp. 221–234.

Fortin, N., Lemieux, T. and Firpo, S. (2011) 'Decomposition Methods in Economics', *Handbook of Labor Economics*, 4(PART A), pp. 1–102. doi: 10.1016/S0169-7218(11)00407-2.

Foy, P. *et al.* (2016) 'Reviewing the PIRLS 2016 Achievement Item Statistics', *Pirls 2016*, pp. 1–26. Available at: https://timssandpirls.bc.edu/publications/pirls/2016-methods/P16_MP_Chap10_Review_Item_Stats.pdf.

Francis, B. *et al.* (2017) 'Attainment Grouping as self-fulfilling prophesy? A mixed methods exploration of self confidence and set level among Year 7 students', *International Journal of Educational Research*. Elsevier, 86(October), pp. 96–108. doi: 10.1016/j.ijer.2017.09.001.

Gamboa, L. F., Rodríguez Acosta, M. and García-Suaza, A. (2013) 'Diferencias en las motivaciones y el rendimiento académico', *Lecturas de Economía*, (78), pp. 9–44.

Garcia, E. (2014) 'The Need To Address Noncognitive Skills in the Education Policy', *Economic Policy Institute Briefing Paper #386*, pp. 1–36. Available at: Retrieved from www.epi.org, May 1, 2015.

Garcia, M. E. G. (2013) *What we learn in school: Cognitive and non-cognitive skills in the educational production function*, Doctoral Dissertation. Columbia University. Available at: http://ezphost.dur.ac.uk/login?url=https://search.proquest.com/docview/1357123846?accountid=14533%0Ahttp://openurl.ac.uk/ukfed:dur.ac.uk?genre=dissertations+%26+theses&issn=&title=What+we+learn+in+school%3A+Cognitive+and+non-cognitive+skills+in+the+educat.

Gierl, M. J. *et al.* (2003) 'Identifying Content and Cognitive Skills that Produce Gender Differences in Mathematics: A Demonstration of the Multidimensionality-Based DIF Analysis Paradigm', *Journal of Educational Measurement*, 40(4), pp. 281–306. doi: 10.1111/j.1745-3984.2003.tb01148.x.

Goodenow, C. (1993) 'Classroom Belonging among Early Adolescent Students: Relationships to Motivation and Achievement', *The Journal of Early Adolescence*, 13(1), pp. 21–43. doi: 10.1177/0272431693013001002.

Grant, L. (2014) 'How do South Africa's children travel to school?', *Mail and Guardian*, 22 July. [Online]. Accessed June 2019. Available: https://mg.co.za/data/2014-07-22-how-do-south-africas-children-travel-to-school

Grant, M. J. and Behrman, J. R. (2010) 'Gender gaps in educational attainment in less developed countries', *Population and Development Review*, 36(1), pp. 71–89. doi: 10.1111/j.1728-4457.2010.00318.x.

Gustafsson, M. (2020) A revised PIRLS 2011 to 2016 trend for South Africa and the importance of analysing the underlying microdata. *Stellenbosch Economics Working Papers*, No. 02/2020. Stellenbosch.

Gustafsson, M. and Nuga, C. (2020) How is the COVID-19 pandemic affecting educational quality in South Africa? Evidence to date and future risks. *NIDS-CRAM Working Papers.*

Halle, T. G. and Darling-Churchill, K. E. (2016) 'Review of measures of social and emotional development', *Journal of Applied Developmental Psychology*. The Authors, 45, pp. 8–18. doi: 10.1016/j.appdev.2016.02.003.

Hanushek, E. A. (1979) 'Conceptual and Empirical Issues in the Estimation of Educational Production Functions', *Journal of Human Resources*, 14(3), pp. 351–388.

Hanushek, E. A. (2021) 'Addressing cross-national generalizability in educational impact evaluation', *International Journal of Educational Development*. Elsevier Ltd, 80(December 2020), p. 102318. doi: 10.1016/j.ijedudev.2020.102318.

Hanushek, E. A., Machin, S. and Woessmann, L. (2011) *Handbook of the Economics of Education, Volume 4*. 1st edn. Amsterdam: North Holland.

HCEO. n.d. Human Capital and Economic Opportunity Global Working Group: About. [Online]. Accessed September 2020. Available: https://hceconomics.uchicago.edu/about

He, X. *et al.* (2021) 'IQ, grit, and academic achievement: Evidence from rural China', *International Journal of Educational Development*. Elsevier Ltd, 80(October 2020), p. 102306. doi: 10.1016/j.ijedudev.2020.102306.

Heckman, J. J. (2000) 'Policies to foster human capital', *Research in Economics*, 54(December 1998), pp. 3–56. doi: 10.1006/reec.1999.0225.

Heckman, J. J. (2006) 'Skill Formation and the Economics of Investing in Disadvantaged Children', *Science*, 312(5782), pp. 1900–1902.

Heckman, J. J. and Corbin, C. O. (2016) 'Capabilities and Skills', *Journal of Human Development and Capabilities*. Taylor & Francis, 17(3), pp. 342–359. doi: 10.1080/19452829.2016.1200541.

Heckman, J. J. and Kautz, T. (2012) *Hard Evidence on Soft Skills*, *IZA Discussion Paper*, No. 6580. Bonn, Germany.

Hemmings, B., Grootenboer, P. and Kay, R. (2011) 'Predicting mathematics achievement: The influence of prior achievement and attitudes', *International Journal of Science and Mathematics Education*, 9(3), pp. 691–705. doi: 10.1007/s10763-010-9224-5.

Hoadley, U. (2018) *'Pedagogy in Poverty: Lessons from twenty years of curriculum reform in South Africa'*. Routledge: New York.

House, J. D. (2003) 'Self-beliefs and reading achievement of elementary-school students in Hong Kong and the united states: results from the PIRLS 2001 assessment', *Child Study Journal*, 33(4).

House, J. D. and Telese, J. A. (2016) 'Achievement of Eighth-Grade Students in Korea on the TIMSS 2011 Assessment: Effects of Confidence in Mathematics and Engagement in Mathematics Lessons', *Education*, 137(1), pp. 77–81.

Howie, S. *et al.* (2017) *Pirls Literacy 2016: South African Children's Reading Literacy Achievement*. Pretoria.

Howie, S. J. *et al.* (2016) *PIRLS Literacy 2016: South African Highlights Report*. Pretoria.

Huang, H. and Zhu, H. (2017) 'High Achievers from Low Socioeconomic Backgrounds: The Critical Role of Disciplinary Climate and Grit.', *Mid-Western Educational Researcher*, 29(2), pp. 93–116.

Hughes, J. N., Myung, H. I. and Allee, P. J. (2015) 'Effect of School Belonging Trajectories in Grades 6-8 on Achievement: Gender and Ethnic Differences', *Journal of School Psychology*, 176(1), pp. 139–148. doi: 10.1016/j.physbeh.2017.03.040.

Human Sciences Research Council (2017) *TIMSS Item Diagnostic Report: South Africa - Grade 5: Numeracy*. Cape Town.

Ibourk, A. (2013) 'Determinants of Educational Achievement in Morocco: A Micro-Econometric Analysis Applied to the PIRLS Study', *Regional and Sectoral Economic Studies*, 13(2), pp. 179–190.

Ikeda, M. and García, E. (2014) 'Grade repetition: A comparative study of academic and non-academic consequences', *OECD Journal: Economic Studies*, 2013(1), pp. 269–315. doi: 10.1787/eco_studies-2013-5k3w65mx3hnx.

Ivcevic, Z. and Brackett, M. (2014) 'Predicting school success: Comparing Conscientiousness, Grit, and Emotion Regulation Ability', *Journal of Research in Personality*. Elsevier Inc., 52, pp. 29–36. doi: 10.1016/j.jrp.2014.06.005.

Jacob, B. A. (2002) Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *NBER Working Papers*, No. 8964. Cambridge, MA.

Jann, B. (2008a) 'A Stata implementation of the Blinder-Oaxaca', *ETH Zurich Sociology Working Paper*, 8(5), pp. 453–479.

Jann, B. (2008b) 'The Blinder-Oaxaca decomposition for linear regression models', *Stata Journal*, 8(4), pp. 453–479. doi: 10.1177/1536867x0800800401.

Johnes, G., Johnes, J. and López-Torres, L. (2017) 'Human capital and returns to education', *Handbook of Contemporary Education Economics*, (1961), pp. 1–20. doi: 10.4337/9781785369070.00006.

Jukes, M. *et al.* (2018) Developing a Culturally Relevant Assessment of Social and Emotional Learning for Tanzania. *Global Education Review*. 5(2): 160-188.

Kannangara, C. S. *et al.* (2018) 'All that glitters is not grit: Three studies of grit in University Students', *Frontiers in Psychology*, 9(AUG), pp. 1–15. doi: 10.3389/fpsyg.2018.01539.

Kautz, T. *et al.* (2014) Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. *IZA Discussion Papers*, No. 8698. Bonn, Germany.

Kennedy, A. (2008) 'Examining Gender and Fourth Graders ' Reading Habits and Attitudes in PIRLS 2001 and 2006', *IEA International Research Conference*.

Kim, H. Y., Brown, L. and Weiss-Yagoda, J. (2018) *IRC Healing Classrooms Remedial Tutoring Programming Improves Nigerien and Nigerian Children's Learning*. New York.

Kingdon, G. G. (2002) 'The gender gap in educational attainment in India: How much can be explained?', *Journal of Development Studies*, 39(2), pp. 25–53. doi: 10.1080/00220380412331322741.

Kline, P. (2000) *Handbook of Psychological Testing*. 2nd edn. London: Routledge.

Kraft, M. A. (2019) 'Teacher effects on complex cognitive skills and social-emotional competencies', *Journal of Human Resources*, 54(1), pp. 1–36. doi: 10.3368/JHR.54.1.0916.8265R3.

Laajaj, R. and Macours, K. (2017) 'Measuring Skills in Developing Countries', *The World Bank*, (January), pp. 1–80.

Legewie, J. and DiPrete, T. A. (2012) 'School Context and the Gender Gap in Educational Achievement', *American Sociological Review*, 77(3), pp. 463–485. doi: 10.1177/0003122412440802.

Light, A. and Nencka, P. (2019) 'Predicting educational attainment: Does grit compensate for low levels of cognitive ability?', *Learning and Individual Differences*. Elsevier, 70(January), pp. 142–154. doi: 10.1016/j.lindif.2019.02.002.

Logan, S. and Johnston, R. S. (2009) 'Gender differences in reading ability and attitudes: examining where the differences lie', *Research in Reading*, 32(2), pp. 199–214.

Ma, X. (1997) 'Reciprocal relationships between attitude toward mathematics and achievement in mathematics', *Journal of Educational Research*, 90(4), pp. 221–229. doi: 10.1080/00220671.1997.10544576.

Ma, X. and Kishor, N. (2006) 'Assessing the Relationship between Attitude toward Mathematics and Achievement in Mathematics: A Meta-Analysis', *Journal for Research in Mathematics Education*, 28(1), p. 26. doi: 10.2307/749662.

Ma, X. and Xu, J. (2004) 'Determining the Causal Ordering between Attitude toward Mathematics and Achievement in Mathematics', *American Journal of Education*, 110(3), pp. 256–280.

Malanchini, M. *et al.* (2017) 'Reading self-perceived ability, enjoyment and achievement: A genetically informative study of their reciprocal links over time', *Developmental Psychology*, 53(4), pp. 698–712. doi: 10.1037/dev0000209.

Marcenaro–Gutierrez, O., Lopez–Agudo, L. A. and Ropero-García, M. A. (2018) 'Gender Differences in Adolescents' Academic Achievement', *Young*, 26(3), pp. 250–270. doi: 10.1177/1103308817715163.

Marinak, B. A. and Gambrell, L. B. (2010) 'Reading motivation: Exploring the elementary gender gap', *Literacy Research and Instruction*, 49(2), pp. 129–141. doi: 10.1080/19388070902803795.

Marsh, H. W. (1984) 'The bias of negative worded items in ratings scales for preadolescent children: A cognitive-developmental phenomenon', *Developmental Psychology*, 22(1), p. 28.

Marsh, H. W. *et al.* (2013) 'Factorial, convergent, and discriminant validity of timss math and science motivation measures: A comparison of arab and anglo-saxon countries', *Journal of Educational Psychology*, 105(1), pp. 108–128. doi: 10.1037/a0029907.

Mason, H. D. (2018) 'Grit and academic performance among first-year university students: A brief report', *Journal of Psychology in Africa*, 28(1), pp. 66–68. doi: 10.1080/14330237.2017.1409478.

Masten, A. S. and Obradovic, J. (2008) 'Disaster Preparation and Recovery: Lessons from Research on Resilience in Human Development', *Ecology and Society*, 13(1). doi: Artn 9.

McGeown, S. P. *et al.* (2015) 'The relationship between young children's enjoyment of learning to read, reading attitudes, confidence and attainment', *Educational Research*. Routledge, 57(4), pp. 389–402. doi: 10.1080/00131881.2015.1091234.

McKenzie, D. (2016) *Book Review: Grit - Takeaways for Development Economists and Parents*, *Development Impact*. [Online]. Accessed February 2020. Available:

https://blogs.worldbank.org/impactevaluations/book-review-grit-takeaways-development-economists-and-parents

McKown, C. (2019) Student Social and Emotional Competence Assessment: The Current State of the Field and a Vision for Its Future. *Collaborative for Academic, Social, and Emotional Learning.*

Mencarini, L., Pasqua, S. and Romiti, A. (2019) 'Single-mother families and the gender gap in children's time investment and non-cognitive skills', *Review of Economics of the Household*, 17(1), pp. 149–176. doi: 10.1007/s11150-017-9385-x.

Miranda, J. and Domingues, D. (2018) '*Persistent effects of social and emotional skills on students' achievement: Evidence from Brazil*'. [Online]. Accessed April 2020. Available: https://www.fearp.usp.br/images/ppge/Paper_Prof._Daniel_Jéssica_Miranda.pdf

Miscevic-Kadijevic, G. (2015) 'TIMSS 2011: Relationship between Self-Confidence and Cognitive Achievement for Serbia and Slovenia', *Revista Electrónica de Investigación Educativa*, 17(3), pp. 109–115.

Mol, S. E. and Jolles, J. (2014) 'Reading enjoyment amongst non-leisure readers can affect achievement in secondary school', *Frontiers in Psychology*, 5(OCT), pp. 1–10. doi: 10.3389/fpsyg.2014.01214.

Muenks, K. *et al.* (2017) 'How True Is Grit? Assessing Its Relations to High School and College Students' Personality Characteristics, Self-Regulation, Engagement, and Achievement', *Journal of Educational Psychology*, 109(5), pp. 599–620. doi: 10.1037/edu0000153.supp.

Mullis, I. V. S. *et al.* (2015) *Timss 2015 International Results in Mathematics*. Boston. Available at: http://timss2015.org/timss-2015/science/student-achievement/distribution-of-science-achievement/.

Mullis, I. V. S. *et al.* (2016) *TIMSS 2015 International Results in Mathematics*, *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. Paris. doi: 10.4135/9781506326139.n704.

Mullis, I. V. S. *et al.* (2017) *PIRLS 2016 International Results in Reading*. Available at: http://timssandpirls.bc.edu/pirls2016/international-results/.

OECD (2011) *Against the Odds: Disadvantaged Students Who Succeed in School*. OECD Publishing.

OECD (2015) *Skills for Social Progress: The Power of Social and Emotional Skills*. Paris: OECD Publishing.

Osher, D. *et al.* (2017) 'Advancing the Science and Practice of Social and Emotional Learning', *Review of Research in Education*, 40(1), pp. 644–681. doi: 10.3102/0091732x16673595.

Osher, D. *et al.* (2018) 'Drivers of human development: How relationships and context shape learning and development', *Applied Developmental Science*. Taylor & Francis, pp. 1–31. doi: 10.1080/10888691.2017.1398650.

Park, Y. (2011) 'How motivational constructs interact to predict elementary students' reading performance: Examples from attitudes and self-concept in reading', *Learning and Individual Differences*, 21(4), pp. 347–358.

Perkins-Gough, D. (2013) 'The Significance of Grit: A Conversation with Angela Lee Duckworth', *Educational Leadership*, 71(1), pp. 14–20.

Petscher, Y. (2010) 'A meta-analysis of the relationship between student attitudes towards reading and achievement in reading', *Journal of Research in Reading*, 33(4), pp. 335–355. doi: 10.1111/j.1467-9817.2009.01418.x.

Ponzo, M. (2012) Does bullying reduce educational achievement? An evaluation using matching estimators, *Economic Policy Discussion Papers*, No. 36064. doi: 10.1227/01.NEU.0000349921.14519.2A.

Popp, A. M. *et al.* (2014) 'Gender, bullying victimization, and education', *Violence and Victims*, 29(5), pp. 843–856. doi: 10.1891/0886-6708.VV-D-13-00047.

Poropat, A. E. (2009) 'A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance', *Psychological Bulletin*, 135(2), pp. 322–338. doi: 10.1037/a0014996.

Pritchett, L. and Sandefur, J. (2013) 'Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix', *Centre for Global Development Working Papers,* No. 336. doi: 10.2139/ssrn.2364580.

Puerta, M. L. S. and Valerio, A. (2016) 'Definitions: What are socio-emotional skills', in *Taking Stock of Programs to Develop Socioemotional Skills: A Systematic Review of Program Evidence*. World Bank Group, pp. 15–23.

Ravallion, M. (2020) *Should The Randomistas (Continue to) Rule?*, *NBER Working Papers*. doi: 10.1088/1751-8113/44/8/085201.

Retelsdorf, J., Köller, O. and Möller, J. (2014) 'Reading achievement and reading self-concept - Testing the reciprocal effects model', *Learning and Instruction*. Elsevier Ltd, 29, pp. 21–30. doi: 10.1016/j.learninstruc.2013.07.004.

Reynolds, A. J. and Walberg, H. J. (1992) 'A Process Model of Mathematics Achievement and Attitude', *Journal for Research in Mathematics Education*, 23(4), pp. 306–328.

Rimfeld, K. *et al.* (2016) 'True grit and genetics: Predicting academic achievement from personality', *Journal of Personality and Social Psychology*, 111(5), pp. 780–789. doi: 10.1037/pspp0000089.

Ris, E. (2016) 'The problem with teaching "grit" to poor kids? They already have it. Here's what they really need.', *The Washington Post*. [Online]. Accessed January 2020. Available: https://www.washingtonpost.com/news/answer-sheet/wp/2016/05/10/the-problem-with-teaching-grit-to-poor-kids-they-already-have-it-heres-what-they-really-need/?noredirect=on&utm_term=.fd53d92f73cc.

Ross, C. E. and Mirowsky, J. (2006) 'Sex differences in the effect of education on depression: Resource multiplication or resource substitution?', *Social Science and Medicine*, 63(5), pp. 1400–1413. doi: 10.1016/j.socscimed.2006.03.013.

Rosselli, M. *et al.* (2009) 'Gender differences and cognitive correlates of mathematical skills in school-aged children', *Child Neuropsychology*, 15(3), pp. 216–231. doi: 10.1080/09297040802195205.

Roszkowski, M. J. and Soven, M. (2010) 'Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire', *Assessment and Evaluation in Higher Education*, pp. 117–134. doi: 10.1080/02602930802618344.

Salazar, M. S. (2015) 'The dilemma of combining positive and negative items in scales', *Psicothema*, 27(2), pp. 192–199. doi: 10.7334/psicothema2014.266.

Sampson, R. J. (2016) 'The Characterological Imperative: On Heckman, Humphries, and Kautz's 'The Myth of Achievement Tests: The GED and the Role of Character in American Life', *Journal of Economic Literature*, 54(2), pp. 493–513.

Sánchez, B., Colón, Y. and Esparza, P. (2005) 'The role of sense of school belonging and gender in the academic adjustment of Latino adolescents', *Journal of Youth and Adolescence*, 34(6), pp. 619–628. doi: 10.1007/s10964-005-8950-4.

Sandoval-Hernandez, A. and Bialowoski, P. (2016) 'Factors and conditions promoting academic resilience: a TIMSS-based analysis of five Asian education systems', *Asia Pacific Education Review*, 17, pp. 511–520. doi: 10.1007/s12564-016-9447-4.

Sandoval-Hernandez, A. and Cortes, D. (2012) *Factors and conditions that promote academic resilience: A cross-country perspective*. [Online]. Accessed March 2019. Available: https://www.academia.edu/download/7085233/1793097_ICSEI_2012.pdf

Sari, M. (2012) 'Sense of School Belonging Among Elementary School Students', *Çukurova University Faculty of Education Journal*, 41(1), pp. 1–11.

Sayed, Y. (2002) ‘Changing forms of teacher education in South Africa: A case study of policy change’, *International Journal of Educational Development*, 22(3–4), pp. 381–395. doi: 10.1016/S0738-0593(01)00062-1.

Scheithauer, H. *et al.* (2006) ‘Physical, Verbal, and Relational Forms of Bullying Among German Students: Age Trends, Gender Differences, and Correlates’, *Aggressive Behaviour*, 32(3), pp. 187–194. doi: 10.1002/ab.

Scorza, P. *et al.* (2017) ‘Towards clarity in research on “non-cognitive” skills: Linking executive functions, self-regulation, and economic development to advance life outcomes for children, adolescents and youth globally’, *Human Development*, 58(6), pp. 313–317. doi: 10.1159/000443711.

Shepherd, D. (2011) 'Constraints to school effectiveness: what prevents poor schools from delivering results*?' Stellenbosch Economics Working Papers*. Stellenbosch.

Shepherd, D. (2013) ‘A question of efficiency: decomposing South African reading test scores using PIRLS 2006’, *Stellenbosch Economics Working Papers*. Available at: http://ideas.repec.org/p/sza/wpaper/wpapers196.html.

Singh, K., Granville, M. and Dika, S. (2002) ‘Mathematics and science achievement: Effects of motivation, interest, and academic engagement’, *Journal of Educational Research*, 95(6), pp. 323–332. doi: 10.1080/00220670209596607.

Spaull, N. (2013) ‘Poverty & privilege: Primary school inequality in South Africa’, *International Journal of Educational Development*. Elsevier Ltd, 33(5), pp. 436–447. doi: 10.1016/j.ijedudev.2012.09.009.

Spaull, N. (2014) *Education quality in South Africa and sub-Saharan Africa: An Economic Approach*. Doctoral Thesis. Stellenbosch University.

Spaull, N. (2019) ‘Equity: A Price Too High to Pay?’, in Spaull, N. and Jansen, J. (eds) *South African Schooling: The Enigma of Inequality*. 1st edn. Zurich: Springer Switzerland, pp. 1–24.

Spaull, N. and Kotze, J. (2015) ‘Starting behind and staying behind in South Africa: The case of insurmountable learning deficits in mathematics’, *International Journal of Educational Development*. Elsevier Ltd, 41, pp. 13–24. doi: 10.1016/j.ijedudev.2015.01.002.

Spaull, N. and Makaluza, N. (2019) ‘Girls Do Better: The pro-female gender gap in learning outcomes in South Africa’, *Agenda*. 33(4), pp. 207–224. doi: 10.1080/10130950.2019.1672568.

Spaull, N. and Pretorius, E. (2019) ‘Still Falling at the First Hurdle: Examining Early Grade Reading in South Africa’, *South African Schooling: The Enigma of Inequality*, pp. 147–168. doi: 10.1007/978-3-030-18811-5_8.

Spaull, N., Pretorius, E. J. and Mohohlwane, N. (2018) *Investigating the Comprehension Iceberg: Developing empirical benchmarks for early grade reading in agglutinating African languages*.

Spinath, B., Eckert, C. and Steinmayr, R. (2014) 'Gender differences in school success: What are the roles of students' intelligence, personality and motivation?', *Educational Research*, 56(2), pp. 230–243. doi: 10.1080/00131881.2014.898917.

Stankov, L. and Lee, J. (2014) 'Quest for the best non-cognitive predictor of academic achievement', *Educational Psychology*, 34(1), pp. 1–8.

Steele, C. M. (1997) 'A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance', *American Psychologist*, 52(6), pp. 613–629.

Steinmayr, R. and Spinath, B. (2008) 'Sex Differences in School Achievement: What Are the Roles of Personality and Acievement Motivation', *European Journal of Personality*, 22, pp. 185–209. doi: 10.1002/per.

Strand, S. (2010) 'Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement', *School Effectiveness and School Improvement*, 21(3), pp. 289–314. doi: 10.1080/09243451003732651.

Sunny, B. S. *et al.* (2017) 'Failing to progress or progressing to fail? Age-for-grade heterogeneity and grade repetition in primary schools in Karonga district, northern Malawi', *International Journal of Educational Development*. Elsevier Ltd, 52, pp. 68–80. doi: 10.1016/j.ijedudev.2016.10.004.

Tapia, M. and Marsh, G. E. (2000) 'Effect of Gender, Achievement in Mathematics, and Ethnicity on Attitudes Toward Mathematics'. [Online]. Acccessed April 2019. Available: https://eric.ed.gov/?id=ED449044.

Taylor, S. and Yu, D. (2009) *The importance of socio-economic status in determining educational achievement in South Africa*, *Stellenbosch Economic Working Papers*, No. 01–2009. Stellenbosch.

Thomson, S. *et al.* (2017) *PIRLS 2016: Reporting Australia's Results*. Victoria.

Topçu, M. S., Erbilgin, E. and Arikan, S. (2016) 'Factors predicting Turkish and Korean students' science and mathematics achievement in timss 2011', *Eurasia Journal of Mathematics, Science and Technology Education*, 12(7), pp. 1711–1737. doi: 10.12973/eurasia.2016.1530a.

Tough, P. (2011) *What if the Secret to Success in Life is Failure?*, *The New York Times Magazine*. [Online]. Accessed February 2020. Available: http://www.ocmboces.org/tfiles/folder1626/HSP%20Agenda%20%26%20Reading%2010-5-11.pdf

Tough, P. (2016) *How Kids Learn Resilience*, *The Atlantic*. [Online]. Accessed February 2020. Available:

http://simplifymy.s3.amazonaws.com/c%2Ftitlei%2Fsched%2Ffiles%2Fhandouts%2FUpload.only-How%20Kids%20Learn%20Resilience.pdf

Tovar García, E. D. (2017) 'The impact of perseverance and passion for long term goals (GRIT) on educational achievements of migrant children: Evidence from Tatarstan, Russia', *Psicologia Educativa*. Colegio Oficial de Psicólogos de Madrid, 23(1), pp. 19–27. doi: 10.1016/j.pse.2017.02.003.

Tse, S. K. and Xiao, X. yun (2014) 'Differential influences of affective factors and contextual factors on high-proficiency readers and low-proficiency readers: a multilevel analysis of PIRLS data from Hong Kong', *Large-Scale Assessments in Education*, 2(1), pp. 1–24. doi: 10.1186/s40536-014-0006-3.

Twist, L., Schagen, I. and Hodgson, C. (2007) *Readers and Reading: the National Report for England 2006 (PIRLS: Progress in International Reading Literacy Study)*. Slough. Available at: http://eric.ed.gov/?id=ED504613.

UNICEF (no date) *Primary Education*. Available at: https://data.unicef.org/topic/education/primary-education/ (Accessed: 16 May 2019).

Van Broekhuizen, H. and Spaull, N. (2017) The 'Martha Effect': The compounding female advantage in South African higher education. *Stellebosch Economics Working Papers*, No. 142017. Stellenbosch.

Van der Berg, S. (2007) 'Apartheid's enduring legacy: Inequalities in education', *Journal of African Economies*, 16(October 2005), pp. 849–880. doi: 10.1093/jae/ejm017.

Van der Berg, S. *et al.* (2016) *Identifying Binding Constraints in Education*. Stellenbosch.

Van der Berg, S. *et al.* (2019) 'The cost of repetition in South Africa', *Stellenbosch Economics Working Papers*, pp. 117–121. doi: 10.1016/0360-5442(87)90115-0.

Van der Berg, S. (2008) 'How effective are poor schools? Poverty and educational outcomes in South Africa', *Studies in Educational Evaluation*, 34(3), pp. 145–154. doi: 10.1016/j.stueduc.2008.07.005.

Van der Berg, S. and Gustafsson, M. (2019) 'Educational outcomes in post-apartheid South Africa: Signs of progress despite great inequality', in Spaull, N. and Jansen, J. (eds) *South African Schooling: The Enigma of Inequality*. Springer.

Van der Berg, S. and Hofmeyr, H. (2018) *Education in South Africa: Background note to the Systematic Country Diagnostic*. Washington, D.C. Available at: https://openknowledge.worldbank.org/bitstream/handle/10986/29793/WBG-South-Africa-Systematic-Country-Diagnostic-FINAL-for-board-SECPO-Edit-05032018.pdf?sequence=1&isAllowed=y.

Van der Berg, S. and Louw, M. (2006) 'Lessons learnt from SACMEQII: South African student performance in regional context', in. Johannesburg: Investment Choices for Education in Africa.

Van de gaer, E. *et al.* (2009) ‘School Engagement and Language Achievement: A Longitudinal Study of Gender Differences across Secondary School’, *Merill-Palmer Quarterly*, 55(4), pp. 373–405.

Van de gaer, E. *et al.* (2012) ‘The reference group effect: An explanation of the paradoxical relationship between academic achievement and self-confidence across countries’, *Journal of Cross-Cultural Psychology*, 43(8), pp. 1205–1228. doi: 10.1177/0022022111428083.

Vera, G. G., Valenzuela, J. P. and Sotomayor, C. (2015) ‘Against All Odds: Outstanding Reading Performance among Chilean Youth in Vulnerable Conditions’, *Comparative Education Review*, 59(4), pp. 693–716. doi: 10.1086/683108.

Wang, J. H. and Guthrie, J. T. (2004) ‘Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between U.S. and Chinese students’, *Reading Research Quarterly*, 39(2), pp. 162–186.

Wang, J. and Lin, E. (2008) ‘An alternative interpretation of the relationship between self-concept and mathematics achievement: Comparison of Chinese and UD students as a context’, *Evaluation & Research in Education*, 21(3), pp. 154–174.

Weems, G. H. *et al.* (2003) ‘Characteristics of respondents who respond differently to positively and negatively Worded items on Rating Scales’, *Assessment and Evaluation in Higher Education*, 28(6), pp. 587–607. doi: 10.1080/0260293032000130234.

Weems, G. H., Onwuegbuzie, A. J. and Collins, K. M. T. (2009) ‘The Role of Reading Comprehension in Responses to Positively and Negatively Worded Items on Rating Scales’, *Evaluation & Research in Education*, pp. 3–20. doi: 10.1080/09500790608668322.

Weems, G. H., Onwuegbuzie, A. J. and Lustig, D. (2009) ‘Profiles of Respondents Who Respond Inconsistently to Positively- and Negatively-worded Items on Rating Scales’, *Evaluation & Research in Education*, 17(1), pp. 45–60. doi: 10.1080/14664200308668290.

West, M. R. *et al.* (2016) ‘Promise and Paradox: Measuring Students Non-Cognitive Skills and the Impact of Schooling’, *Educational Evaluation and Policy Analysis*, 38(1), pp. 148–170. doi: 10.3102/0162373715597298.

Wills, G. (2017) 'What do you mean by ‘good’? The search for exceptional primary schools in South Africa’s no-fee school system'. *Stellenbosch Economics Working Papers*. Stellenbosch.

Wills, G. and Hofmeyr, H. (2019) ‘Academic resilience in challenging contexts: Evidence from township and rural primary schools in South Africa’, *International Journal of Educational Research*. Elsevier, 98(June), pp. 192–205. doi: 10.1016/j.ijer.2019.08.001.

Wilsenach, C. and Makaure, P. (2018) ‘Gender effects on phonological processing and reading development in Northern Sotho children learning to read in English: A case study of Grade 3 learners’, *South African Journal of Childhood Education*, 8(1), pp. 1–12. doi: 10.4102/sajce.v8i1.546.

World Bank Group (2018) *Learning to Realise Education’s Promise*. Washington, D.C.

World Bank Group (2020) ‘The COVID-19 Pandemic: Shocks to Education and Policy Responses’, *World Bank Group*, 29(8), p. 44. doi: 10.12968/bjon.2020.29.8.456.

Van Wyk, C. (2015) *An overview of Education data in South Africa : an inventory approach*. *Stellenbosch Economics Working Papers*. Stellenbosch.

Zuze, T. L. (2015) ‘A riddle explained: Gender disparities in East African education’, *African Review of Economics and Finance*, 7(2), pp. 29–59.

Zuze, T. L. *et al.* (2017) *TIMSS 2015 Grade 9 National Report: Understanding Mathematics and Science Achievement Amongst Grade 9 Learners in South Africa*. Cape Town.

Zuze, T. L. and Beku, U. (2019) ‘Gender Inequalities in South African Schools: New Complexities’, in Spaull, N. and Jansen, J. D. (eds) *South African Schooling: The Enigma of Inequality*. Cape Town: Springer, pp. 225–241.

APPENDIX A

1. CONSTRUCTION AND CHARACTERISTICS OF SES INDEX

Principal components analysis (PCA) was used to construct an SES index. The index includes variables indicating the presence of six items in the home, parental education and a dummy indicating whether the student attends a school located in a township or remote rural area. The home possession variables are: A computer or tablet at home; a study desk; own cell phone; a gaming console (e.g. Play Station); own bedroom; and internet at home.

Parental education was included in the SES index since education attainment of parents is considered a good proxy for SES. This variable was included as a dummy indicating whether parents have completed high school. The logic behind this decision was as follows: Parental education is collected as an ordinal categorical variable in the PIRLS data. In principal components analysis, ordinal categorical variables are treated as continuous variables, which creates problems for interpretation of coefficients on this index. One solution to this is to treat each category as a dummy variable, but these dummy variables will necessarily be correlated with each other, which will "confuse" the principal components analysis and result in an index that does not accurately capture SES (Taylor & Yu, 2009). While there are other methods for dealing with this problem, it was decided that including parental education as a dummy indicating high school completion provided the simplest solution, since analyses of labour market earnings in South Africa show large returns to high school graduation. It is thus expected that having a parent who has completed matric should contribute significantly to SES. In addition, including parental education as an ordinal categorical variable does not add to the correlation coefficient any more than adding parental education as a dummy indicating completion of high school does.

2. IMPUTATION OF MISSING DATA

(i) Home possessions

Following Taylor & Yu (2009), missing data on the home possessions variables was dealt with by imputation on the assumption that students who did not provide an answer did not have access to the relevant possession at home. Zero imputation is justified when one can reasonably expect an unwillingness to give an answer to be more common amongst students who do not possess a particular item at home (Taylor & Yu, 2009: 16). In addition, missingness on the home possessions variables is negatively correlated with both reading scores and parents' education. This suggests failure to answer the home possessions questions was more likely among students of low SES.

(ii) Parental education

Missing information on mother and father's education was imputed as the mean of these variables at the school level. This consideration was primarily informed by the fact that the South African PIRLS data shows little variation in educational attainment of parents at the school level. There is evidence that this is an appropriate method of dealing with missing information on parents' education: Proportions of parents who completed high school in this imputed dataset are very similar to estimates of this measure based on other nationally representative datasets. For example, according to estimates based on the National Income Dynamics Study (NIDS), between 42% and 44% of adults in the relevant age group have completed matric (Branson & Lam, 2009). The proportions of students whose mothers and fathers have completed matric are 44% and 43%, respectively, according to the imputed PIRLS data. This provides evidence that this method of imputing missing values for the parental education variables results in imputed data that is a fairly accurate reflection of parental education in the PIRLS sample.

Unfortunately, there are a number of instances in the South African PIRLS data where all students in a school have missing values for the parental education variables. Naturally mean parental education at the school level could not be used to impute these missing values, and these students were dropped from the analysis. This resulted in a loss of 1359 observations from the original sample of 12 810 (10.6%).

Table A1: Explanatory variables used in the multivariate analysis: PIRLS 2016 (Grade 4)

Variable	Questionnaire	Description	Response categories and codes
Individual characteristics			
Female	Student	Gender: Female	
Age	Student	Age in years	
Absent	Student	About how often are you absent from school?	1: Never or almost never 2: Once a month 3: Once every two weeks 4: Once a week
How often is the test language spoken at home?	Student		1: Never 2: Sometimes 3: Almost always 4: Always
Student attitudes			
Bullying index	Student	During this school year, how often have other students from your school done any of the following things to you? Made fun of me or called me names Left me out of their games or activities Spread lies about me Stole something from me Hit or hurt me Made me do things I didn't want to do Shared embarrassing information about me Threatened me	1: Almost never 2: About monthly 3: About weekly See http://timssandpirls.bc.edu/pirls2016/international-results/pirls/school-safety/student-bullying/ for a description of how student responses were combined into an overall bullying index.
Sense of school belonging (index)	Student	What do you think about your school? Tell how much you agree with these statements: I like being in school I feel safe when I am at school I feel like I belong at this school Teachers at my school are fair to me I am proud to go to this school	Four-point Likert scale from "Agree a lot" to "Disagree a lot". See http://timssandpirls.bc.edu/pirls2016/international-results/pirls/school-climate/student-sense-of-belonging/ for a description of how student responses were combined into an overall index indicating students' sense of school belonging.
Reading enjoyment (index)	Student	What do you think about reading? Tell how much you agree with each of these statements: I like talking about what I read with other people I would be happy if someone gave me a book as a present Reading is boring* I would like to have more time for reading I enjoy reading I learn a lot from reading I like to read things that make me think I like when a book helps me imagine other worlds How often do you do these things outside of school? I read for fun I read to find out about things I want to learn *Reverse coded	See http://timssandpirls.bc.edu/pirls2016/international-results/pirls/student-engagement-and-attitudes/students-like-reading/ for a description of how student responses were combined into an overall index indicating how much students enjoy reading.

Table A1: Explanatory variables used in the multivariate analysis: PIRLS (Grade 4) (Cont.)

Variable	Questionnaire	Description	Response categories and codes
Reading engagement (index)	Student	How much do you agree with these statements about your reading lessons: I like what I read about in school My teacher gives me interesting things to read I know what my teacher expects me to do My teacher is easy to understand I am interested in what my teacher says My teacher encourages me to say what I think about what I have read My teacher lets me show what I have learned My teacher does a variety of things to help us learn My teacher tells me how to do better when I make a mistake”	Four-point Likert scale from “Agree a lot” to “Disagree a lot”. See http://timssandpirls.bc.edu/pirls2016/international-results/pirls/student-engagement-and-attitudes/students-engaged-in-reading-lessons/ for a description of how student responses were combined into an overall index indicating student engagement in reading lessons.
Confidence in reading (index)	Student	How well do you read? Tell how much you agree with each of these statements: I usually do well in reading Reading is easy for me I have trouble reading stories with difficult words* Reading is harder for me than many of my classmates* Reading is harder for me than any other subject* I am just not good at reading* *Reverse coded	Four-point Likert scale from “Agree a lot” to “Disagree a lot”. See http://timssandpirls.bc.edu/pirls2016/international-results/pirls/student-engagement-and-attitudes/students-confident-in-reading/ for a description of how student responses were combined into an overall index indicating student confidence in reading.
English or Afrikaans school		Dummy indicating whether test was written in English or Afrikaans	0: Test was written in isiNdebele, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, Siswati, Tshivenda, or Xitsonga 1: Test was written in English or Afrikaans
School library	Teacher	Dummy indicating the presence of a library at the school	
Teacher has at least a Bachelor’s degree	Teacher	Dummy indicating whether the teacher has obtained at least a Bachelor’s degree	
SES of school body	Student (derived)	Mean of student SES at the school level (standardised)	
School has at least one computer	Principal	Dummy indicating the school has at least one computer	

Table A2: Explanatory variables used in multivariate analysis: TIMSS 2015 (Grade 9)

Variable	Questionnaire	Description	Response categories and codes
Individual characteristics			
Female	Student	Gender: Female	
Age	Student	Age in years	
Absent	Student	About how often are you absent from school?	1: Never or almost never 2: Once a month 3: Once every two weeks 4: Once a week
How often is the test language spoken at home?	Student		1: Never 2: Sometimes 3: Almost always 4: Always
Student attitudes			
Bullying index	Student	During this school year, how often have other students from your school done any of the following things to you? Made fun of me or called me names Left me out of their games or activities Spread lies about me Stole something from me Hit or hurt me Made me do things I didn't want to do Shared embarrassing information about me Shared something embarrassing about me online Threatened me	1: Almost never 2: About monthly 3: About weekly See http://timssandpirls.bc.edu/timss2015/international-results/timss-2015/mathematics/school-safety/student-bullying/ for a description of how student responses were combined into an overall bullying index.
Sense of school belonging (index)	Student	What do you think about your school? Tell how much you agree with these statements: I like being in school I feel safe when I am at school I feel like I belong at this school I like to see my classmates at school Teachers at my school are fair to me I am proud to go to this school I learn a lot in school	Four-point Likert scale from "Agree a lot" to "Disagree a lot". See http://timssandpirls.bc.edu/timss2015/international-results/timss-2015/mathematics/school-climate/students-sense-of-school-belonging/ for a description of how student responses were combined into an overall index indicating students' sense of school belonging.
Enjoyment of mathematics	Student	How much do you agree with these statements about learning mathematics? I enjoy learning mathematics I wish I did not have to study mathematics* Mathematics is boring* I learn many interesting things in mathematics I like mathematics I like any schoolwork that involves numbers I like to solve mathematics problems I look forward to mathematics class Mathematics is one of my favourite subjects *Reverse-coded	Four-point Likert scale from "Agree a lot" to "Disagree a lot". See http://timssandpirls.bc.edu/timss2015/international-results/timss-2015/mathematics/student-engagement-and-attitudes/students-like-learning-mathematics/ for a description of how student responses were combined into an overall index indicating student enjoyment of mathematics lessons.

Table A2: Explanatory variables used in multivariate analysis: TIMSS 2015 (Grade 9) (Cont.)

Variable	Questionnaire	Description	Response categories and codes
Engagement in mathematics lessons	Student	How much do you agree with these statements about your mathematics lessons? I know what my teacher expects me My teacher is easy to understand I am interested in what my teacher says My teacher gives me interesting things to do My teacher has answers to all my questions My teacher is good at explaining mathematics My teacher lets me show what I have learned My teacher does a variety of things to help us learn My teacher helps me to do better when I make a mistake My teacher listens to what I have to say	Four-point Likert scale from "Agree a lot" to "Disagree a lot". See http://timssandpirls.bc.edu/timss2015/international-results/timss-2015/mathematics/student-engagement-and-attitudes/students-views-on-engaging-teaching-in-mathematics/ for a description of how student responses were combined into an overall index indicating student engagement in mathematics lessons.
Confidence in mathematics	Student	How much do you agree with these statements about mathematics? I usually do well in mathematics Mathematics is more difficult for me than any of my classmates* Mathematics is not one of my strengths* I learn things quickly in mathematics Mathematics makes me nervous* I am good at working out difficult mathematics problems My teacher tells me I am good at mathematics Mathematics is harder for me than any other subject*	Four-point Likert scale from "Agree a lot" to "Disagree a lot". See http://timssandpirls.bc.edu/timss2015/international-results/timss-2015/mathematics/student-engagement-and-attitudes/students-confident-in-mathematics/ for a description of how student responses were combined into an overall index indicating students' confidence in mathematics.
School factors			
School library	Teacher	Dummy indicating the presence of a library at the school	
Teacher has at least a Bachelor's degree	Teacher	Dummy indicating whether the teacher has obtained at least a Bachelor's degree	
School has at least one computer	Principal	Dummy indicating the school has at least one computer	

Table A3: Correlation matrix of explanatory variables (PIRLS 2016)

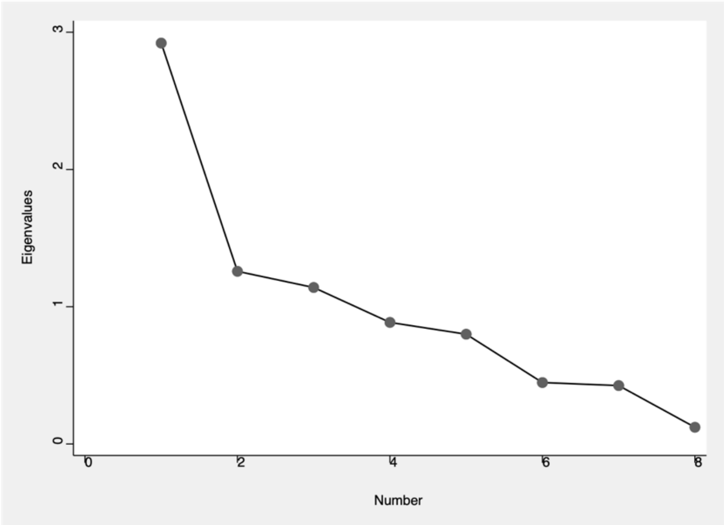
	Female	Age	Often absent	First langu-age	Confi-dence index	Engage-ment index	Enjoy-ment index	Belong-ing index	Bullying index	Library	Com-puters	Teacher has a degree	Peer SES
Female	1.000												
Age	-0.152	1.000											
Often absent	-0.063	0.062	1.000										
First language	-0.010	-0.017	-0.004	1.000									
Confidence index	0.098	-0.095	-0.121	-0.018	1.000								
Engagement index	0.119	-0.095	-0.137	0.084	0.149	1.000							
Enjoyment index	0.146	-0.104	-0.120	0.079	0.200	0.543	1.000						
Belonging index	0.105	-0.073	-0.099	0.105	0.102	0.523	0.411	1.000					
Bullying index	-0.065	0.039	0.096	0.017	-0.092	-0.055	-0.096	-0.034	1.000				
Library	0.003	-0.010	-0.039	-0.029	0.030	0.069	0.039	0.051	0.009	1.000			
Computers	0.013	0.014	-0.006	-0.048	0.020	-0.067	-0.083	-0.065	0.017	0.158	1.000		
Teacher has a degree	0.015	-0.021	-0.017	0.042	0.026	0.048	0.042	0.050	0.013	0.201	0.159	1.000	
Peer SES	0.039	-0.006	-0.090	-0.065	0.120	0.076	0.057	0.013	0.043	0.212	0.194	0.043	1.000

Table A4: Correlation matrix of explanatory variables (TIMSS 2015)

	Female	Age	Often absent	First language	Confidence index	Engagement index	Enjoyment index	Belonging index	Bullying index	Library	Computers	Teacher has a degree	Peer SES
Female	1.000												
Age	-0.199	1.000											
Often absent	-0.035	0.194	1.000										
First language	0.048	-0.051	-0.009	1.000									
Confidence index	-0.061	-0.108	-0.097	-0.017	1.000								
Engagement index	0.054	-0.099	-0.074	0.0000	0.213	1.000							
Enjoyment index	-0.043	-0.098	-0.077	-0.082	0.457	0.438	1.000						
Belonging index	0.048	-0.119	-0.094	-0.064	0.129	0.403	0.298	1.000					
Bullying index	-0.057	0.125	0.143	-0.060	-0.052	-0.112	-0.067	-0.136	1.000				
Library	0.037	-0.055	-0.010	0.078	-0.021	0.025	-0.048	0.018	-0.028	1.000			
Computers	0.021	-0.021	-0.010	0.069	0.002	-0.014	-0.009	0.025	-0.002	0.249	1.000		
Teacher has a degree	0.011	0.037	-0.003	0.026	0.003	0.030	-0.004	0.021	0.015	0.018	-0.011	1.000	
Peer SES	-0.031	-0.152	-0.070	0.081	0.065	0.093	0.022	0.076	-0.029	0.102	0.086	-0.016	1.000

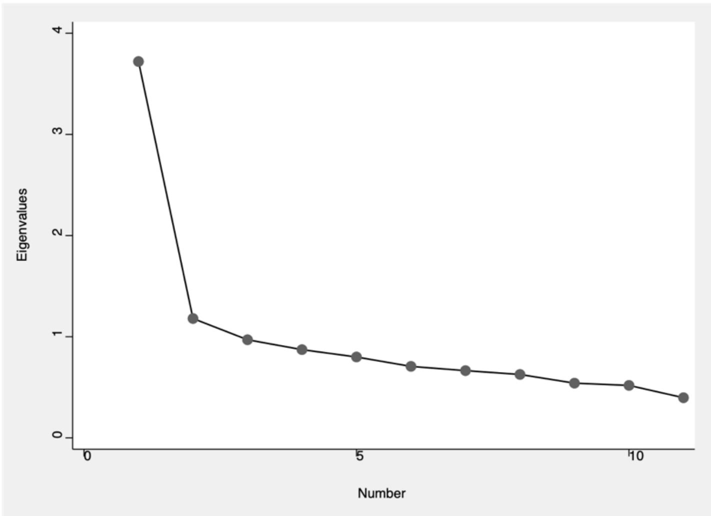
APPENDIX B

Figure B1: Scree plot of Eigenvalues of PCA used to construct school functionality index



Source: Leadership for Literacy

Figure B2: Scree plot of Eigenvalues of PCA used to construct student SES index



Source: Leadership for Literacy

The eight items comprising the adapted Grit-S scale are as follows:

(1) "New ideas sometimes distract me from what I am currently doing."

(2) "Problems and challenges don't discourage me. When I make a mistake I get back up and try again."

(3) "I am sometimes very interested in one thing but only for a short time."

(4) "I work hard to do things well."

(5) "I find it difficult to stick to the same thing."

(6) "I finish whatever I start."

(7) "I can sit still for longer than other children in the class."

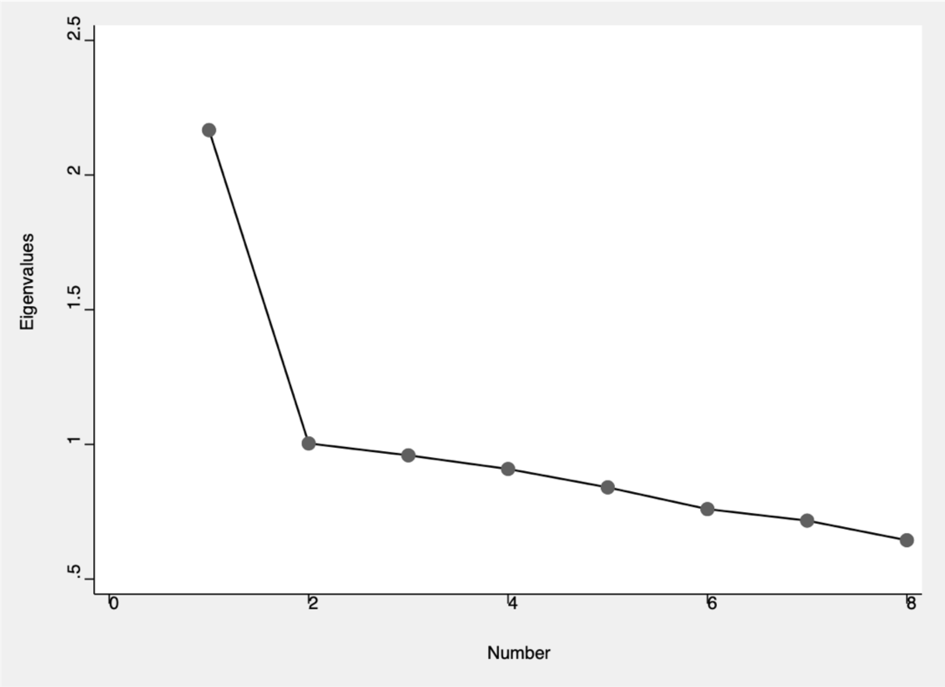
(8) "I do my schoolwork carefully."

Table B1: Results of OLS and HLM regressions estimated for the full Leadership for Literacy Sample

	Full sample: OLS	Full sample: HLM
Perseverance (z-scores)	0.306***	0.283***
	(0.018)	(0.018)
Female	0.256***	0.264***
	(0.036)	(0.034)
Over-age	-0.276***	-0.276***
	(0.048)	(0.046)
Frequency of English use at home	0.172***	0.129***
	(0.037)	(0.037)
Attended Grade R	0.186***	0.153***
	(0.059)	(0.058)
Asset index (z-scores)	0.157***	0.107***
	(0.023)	(0.024)
Asset index^2	0.057***	0.010
	(0.018)	(0.019)
Lives with mother	0.071	0.044
	(0.039)	(0.038)
Lives with father	0.011	0.001
	(0.037)	(0.036)
At least one parent employed	0.136***	0.101**
	(0.047)	(0.045)
School functionality index	0.112***	0.112**
	(0.028)	(0.054)
School uses English as the LOLT in the foundation phase	0.038	0.019
	(0.038)	(0.045)
Peer SES	0.038***	0.050***
	(0.012)	(0.023)
Constant	-1.468***	-1.292***
	(0.134)	(0.176)
R^2	0.31	
N	2,383	2,383

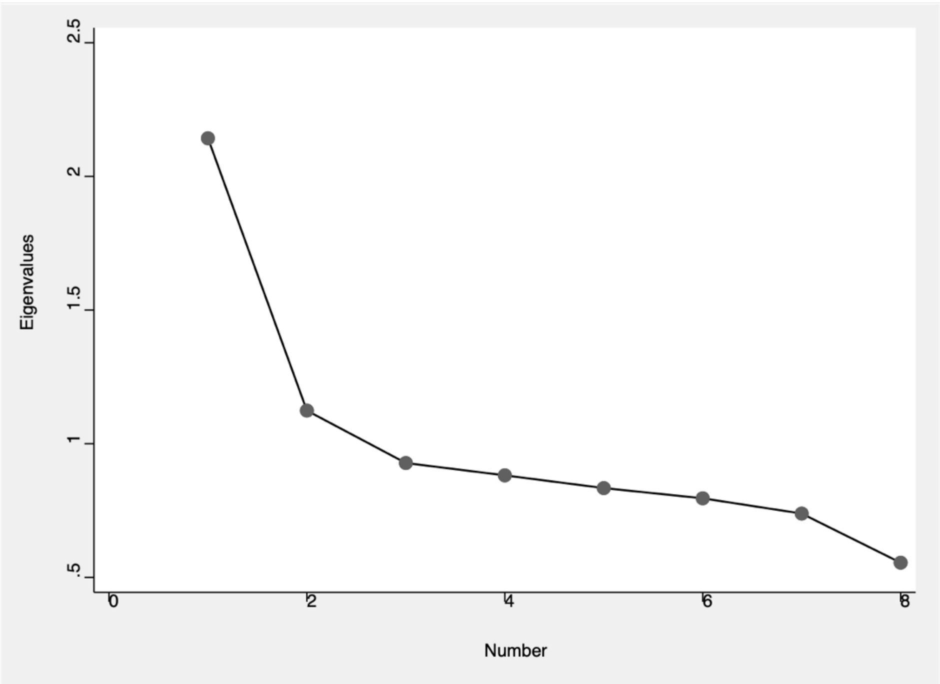
APPENDIX C

Figure C1: Scree plot of Eigenvalues of PCA used to construct asset index (PIRLS)



Source: PIRLS 2016

Figure C2: Scree plot of Eigenvalues of PCA used to construct asset index (TIMSS)



Source: TIMSS 2015

Table C1: Gender differences in student-reported home possessions, by school quintile (PIRLS 2016, Grade 4)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5		Total	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Computer	0.19	0.18	0.32	0.30	0.48	0.42***	0.59	0.56*	0.79	0.76	0.49	0.47***
Study desk	0.32	0.39***	0.55	0.58	0.57	0.61	0.62	0.67**	0.72	0.77**	0.55	0.60***
Cell phone	0.40	0.37	0.54	0.46***	0.63	0.57***	0.69	0.64***	0.75	0.72**	0.61	0.56***
Gaming station	0.20	0.14***	0.32	0.24***	0.43	0.25***	0.54	0.35***	0.72	0.45***	0.45	0.29***
Own room	0.42	0.37**	0.57	0.50***	0.60	0.56**	0.65	0.60**	0.75	0.70***	0.60	0.55***
Internet access	0.12	0.08***	0.25	0.18***	0.29	0.22***	0.42	0.35***	0.54	0.48**	0.33	0.27***
Own books	0.40	0.46***	0.55	0.59***	0.58	0.69***	0.65	0.76***	0.72	0.84***	0.58	0.68***
Daily newspaper	0.24	0.27	0.36	0.36	0.45	0.46	0.49	0.48	0.60	0.65**	0.43	0.45**
N	1,037	987	1,171	1,085	1,322	1,238	1,251	1,252	1,228	1,163	6,009	5,725
Proportion	51%	49%	52%	48%	52%	48%	50%	50%	51%	49%	51%	49%

Notes: Asterisks indicate statistically significant differences in the proportions of boys and girls in each quintile who indicated that they had the given item in their home. Significance levels: *p<0.10; **<p<0.05; ***p<0.01.

Table C2: Gender differences in student-reported home possessions, by school quintile (TIMSS 2015, Grade 5)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5		Total	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Computer	0.33	0.29**	0.42	0.38**	0.42	0.41	0.44	0.43	0.61	0.62	0.49	0.46***
Study desk	0.38	0.37	0.54	0.53	0.52	0.54	0.58	0.63***	0.77	0.78	0.55	0.57
Cell phone	0.47	0.39***	0.54	0.48***	0.63	0.55***	0.68	0.60***	0.76	0.71**	0.61	0.54***
Own room	0.42	0.34***	0.54	0.45***	0.55	0.49***	0.62	0.50***	0.72	0.66***	0.57	0.48***
Internet access	0.22	0.13***	0.28	0.20***	0.32	0.24***	0.41	0.35***	0.66	0.62**	0.37	0.30***
Electricity	0.60	0.61	0.73	0.75	0.81	0.82	0.89	0.91*	0.96	0.96	0.79	0.81
Running water	0.38	0.45***	0.54	0.62***	0.65	0.69**	0.71	0.78***	0.87	0.91**	0.63	0.69***
Fridge	0.74	0.70	0.82	0.81	0.88	0.89	0.93	0.95*	0.97	0.98	0.87	0.86
N	1,063	1,031	1,198	1,125	1,216	1,163	1,225	1,107	899	905	5,601	5,331
Proportion	51%	49%	52%	48%	51%	49%	53%	47%	50%	50%	51%	49%

Notes: Asterisks indicate statistically significant differences in the proportions of boys and girls in each quintile who indicated that they had the given item in their home. Significance levels: *p<0.10; **<p<0.05; ***p<0.01.

Table C3: Responses to student confidence items, by gender and overage (PIRLS)

Questionnaire item	Proportion that answered "Agree a lot"			
	On-track		Overage	
	Boys	Girls	Boys	Girls
I usually do well in reading	0.71***	0.79	0.68*	0.73
Reading is harder for me than for any of my classmates*	0.25	0.30	0.21	0.24
I am just not good at reading*	0.32**	0.39	0.25**	0.31
Reading is easy for me	0.59***	0.69	0.53***	0.63
I have trouble with difficult words*	0.17	0.21	0.14	0.17
Reading is harder for me than any other subject*	0.28**	0.34	0.22	0.24

* Reverse-coded

Table C4: Responses to reading engagement items, by gender and overage (PIRLS)

Questionnaire item	Proportion that answered "Agree a lot"			
	On-track		Overage	
	Boys	Girls	Boys	Girls
I know what my teacher expects me to do	0.62***	0.69	0.58**	0.65
My teacher is easy to understand	0.61***	0.68	0.59*	0.64
I am interested in what my teacher says	0.63***	0.73	0.59**	0.66
My teacher gives me interesting things to read	0.64***	0.74	0.58***	0.68
My teacher lets me show what I have learned	0.61***	0.70	0.58*	0.63
My teacher does a variety of things to help us learn	0.64***	0.74	0.62**	0.68
My teacher tells me how to do better when I make a mistake	0.59***	0.66	0.56**	0.62
My teacher encourages me to say what I think about what I've read	0.60***	0.68	0.58	0.61

Table C5: Responses to reading enjoyment items, by gender and overage (PIRLS)

Questionnaire item	Proportion that answered "Agree a lot"			
	On-track		Overage	
	Boys	Girls	Boys	Girls
I like it when a book helps me to imagine other worlds	0.77	0.79	0.76	0.75
I would be happy if I got a book as a gift	0.23*	0.19	0.32**	0.25
I like reading things that make me think	0.15	0.15	0.25	0.24
I would like to have more time for reading	0.72	0.73	0.66	0.65
I like talking about what I read with other people	0.74*	0.78	0.70	0.70
I think reading is boring*	0.66	0.68	0.64	0.65
I learn a lot from reading	0.65	0.66	0.60	0.61
I like what I read in school	0.65	0.66	0.59	0.60
I enjoy reading	0.69	0.68	0.65	0.64
	Proportion that answered "Every day or almost every day"			
	On-track		Overage	
	Boys	Girls	Boys	Girls
How often do you read for fun?	0.54***	0.63	0.54**	0.60
How often do you read about things you want to learn?	0.62**	0.68	0.60*	0.64

* Reverse-coded

Table C6: Responses to school belonging items, by gender and overage (PIRLS)

Questionnaire item	Proportion that answered "Agree a lot"			
	On-track		Overage	
	Boys	Girls	Boys	Girls
I am proud to go to this school	0.62***	0.72	0.59**	0.66
I feel safe when I am at school	0.63***	0.73	0.59*	0.65
I feel like I belong at this school	0.62*	0.67	0.59*	0.64
Teachers at my school are fair to me	0.54***	0.62	0.52*	0.57
I like being in school	0.75***	0.85	0.72**	0.78

Table C7: Responses to student bullying items, by gender and overage (PIRLS)

Questionnaire item	Proportion that answered "At least once a week"			
	On-track		Overage	
	Boys	Girls	Boys	Girls
How often are you made fun of or called names?	0.34*	0.29	0.36*	0.32
How often have other students left you out of their games or activities?	0.26	0.24	0.37**	0.46
How often have other students spread lies about you?	0.26	0.25	0.27	0.26
How often have other students stolen something from you?	0.27	0.28	0.29	0.29
How often have other students hit you or hurt you?	0.27	0.26	0.28	0.26
How often have other students made you do things you didn't want to do?	0.21	0.19	0.24	0.21
How often have other students shared embarrassing information about you?	0.22	0.22	0.23	0.23
How often have other students threatened you?	0.24	0.24	0.25	0.24

Table C8: Responses to student confidence items, by gender and overage (TIMSS)

Questionnaire item	Proportion that answered "Agree a lot"			
	On-track		Overage	
	Boys	Girls	Boys	Girls
I usually do well in mathematics	0.63*	0.59	0.60**	0.54
Mathematics is harder for me than for any of my classmates*	0.29	0.25	0.34*	0.30
I am just not good at mathematics*	0.29***	0.19	0.29	0.27
I learn things quickly in mathematics	0.62*	0.58	0.53	0.53
Mathematics makes me nervous*	0.30	0.28	0.35	0.33
I am good at working out difficult mathematics problems	0.51	0.48	0.48	0.45
My teacher tells me I am good at mathematics	0.51*	0.47	0.48	0.45
Mathematics is harder for me than any other subject*	0.29*	0.25	0.36*	0.32
Mathematics makes me confused*	0.28	0.25	0.35*	0.31

* Reverse-coded

Table C9: Responses to student engagement items, by gender and overage (TIMSS)

Questionnaire item	Proportion that answered “Agree a lot”			
	On-track		Overage	
	Boys	Girls	Boys	Girls
I know what my teacher expects me to do	0.75	0.76	0.71	0.71
My teacher is easy to understand	0.66	0.69	0.62	0.63
I am interested in what my teacher says	0.70	0.73	0.64*	0.69
My teacher gives me interesting things to do	0.70	0.72	0.65	0.67
My teacher has clear answers to my questions	0.71	0.72	0.66	0.68
My teacher is good at explaining mathematics	0.78	0.80	0.72	0.75
My teacher lets me show what I have learned	0.66	0.69	0.64	0.64
My teacher does a variety of things to help us learn	0.73	0.75	0.66*	0.71
My teacher tells me how to do better when I make a mistake	0.74	0.76	0.67	0.70
My teacher listens to what I have to say	0.70	0.73	0.66	0.69

Table C10: Responses to mathematics enjoyment items, by gender and overage (TIMSS)

Questionnaire item	Proportion that answered “Agree a lot”			
	On-track		Overage	
	Boys	Girls	Boys	Girls
I enjoy learning mathematics	0.77	0.79	0.76	0.75
I wish I did not have to study mathematics*	0.23*	0.19	0.32**	0.25
Mathematics is boring*	0.15	0.15	0.25	0.24
I learn many interesting things in mathematics	0.72	0.73	0.66	0.65
I like mathematics	0.74*	0.78	0.70	0.70
I like any schoolwork that involves numbers	0.66	0.68	0.64	0.65
I like to solve mathematics problems	0.65	0.66	0.60	0.61
I look forward to mathematics lessons	0.65	0.66	0.59	0.60
Mathematics is one of my favourite subjects	0.69	0.68	0.65	0.64

* Reverse-coded

Table C12: Responses to school belonging items, by gender and overage (TIMSS)

Questionnaire item	Proportion that answered “Agree a lot”			
	On-track		Overage	
	Boys	Girls	Boys	Girls
I am proud to go to this school	0.76*	0.80	0.67**	0.74
I feel safe when I am at school	0.73	0.74	0.68*	0.72
I feel like I belong at this school	0.67	0.70	0.62*	0.66
Teachers at my school are fair to me	0.59	0.58	0.53	0.55
I like being in school	0.78*	0.83	0.77*	0.81

Table C13: Responses to student bullying items, by gender and overage (TIMSS)

Questionnaire item	Proportion that answered “At least once a week”			
	On-track		Overage	
	Boys	Girls	Boys	Girls
How often are you made fun of or called names?	0.41***	0.33	0.44**	0.38
How often have other students left you out of their games or activities?	0.29***	0.20	0.30***	0.21
How often have other students spread lies about you?	0.26	0.23	0.28	0.28
How often have other students stolen something from you?	0.32	0.30	0.32	0.30
How often have other students hit you or hurt you?	0.23	0.20	0.25*	0.21
How often have other students made you do things you didn’t want to do?	0.21**	0.16	0.26***	0.18
How often have other students shared embarrassing information about you?	0.22	0.19	0.25	0.22
How often have other students threatened you?	0.23*	0.19	0.24	0.21

Table C14: Results from the first decomposition (includes overage students): PIRLS Grade 4 (Reading)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5	
Gender gap	-0.503***		-0.385***		-0.493***		-0.556***		-0.397***	
	(0.051)		(0.027)		(0.031)		(0.029)		(0.041)	
Boys' average	-0.457***		-0.415***		-0.234***		-0.274***		0.586***	
	(0.062)		(0.045)		(0.047)		(0.050)		(0.100)	
Girls' average	0.046		-0.030		0.259***		0.282***		0.984***	
	(0.054)		(0.029)		(0.034)		(0.043)		(0.089)	
	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained
Overage	-0.036***	0.038	-0.040***	0.035	-0.047***	0.016	-0.038***	-0.040	-0.031***	-0.029**
	(0.010)	(0.019)	(0.008)	(0.022)	(0.009)	(0.019)	(0.006)	(0.022)	(0.008)	(0.011)
Confidence index	-0.045***	-0.017	-0.017***	0.004	-0.046***	-0.002	-0.043***	-0.003	-0.068***	-0.015
	(0.009)	(0.009)	(0.006)	(0.003)	(0.009)	(0.003)	(0.009)	(0.002)	(0.014)	(0.010)
Engagement index	-0.006	-0.002	-0.028***	-0.002	-0.007	-0.000	-0.019***	-0.001	-0.012***	-0.000
	(0.005)	(0.002)	(0.005)	(0.002)	(0.007)	(0.002)	(0.007)	(0.003)	(0.006)	(0.009)
Enjoyment index	-0.022**	0.001	-0.036**	0.001	-0.049**	-0.007	-0.036**	0.001	-0.014	0.002
	(0.009)	(0.003)	(0.007)	(0.000)	(0.012)	(0.007)	(0.006)	(0.004)	(0.009)	(0.003)
Belonging index	0.000	-0.010	-0.014**	-0.001	-0.017	-0.000	-0.006	0.002	-0.004	0.000
	(0.002)	(0.010)	(0.005)	(0.003)	(0.012)	(0.002)	(0.003)	(0.002)	(0.003)	(0.000)
Bullying index	-0.010	-0.004	-0.017***	0.000	-0.013***	-0.003	-0.016***	-0.001	-0.012**	0.003
	(0.006)	(0.010)	(0.005)	(0.001)	(0.003)	(0.004)	(0.005)	(0.001)	(0.005)	(0.006)
Homework	-0.019**	0.009	-0.020***	-0.031	-0.017**	0.008	-0.018***	-0.005	-0.029**	0.040
	(0.007)	(0.017)	(0.004)	(0.016)	(0.006)	(0.024)	(0.005)	(0.014)	(0.012)	(0.035)
Attended ECD	-0.003	0.143	-0.002	-0.060	0.000	0.025	-0.000	-0.061	0.004**	0.095
	(0.002)	(0.106)	(0.002)	(0.076)	(0.002)	(0.119)	(0.000)	(0.077)	(0.002)	(0.108)
Asset index	0.001	-0.056	0.000	-0.006	-0.003	0.001	0.007**	-0.001	0.003	-0.007
	(0.005)	(0.028)	(0.001)	(0.009)	(0.003)	(0.002)	(0.003)	(0.009)	(0.002)	(0.023)
First language	0.007	-0.228***	0.005***	-0.052	-0.000	0.144**	-0.009**	-0.046	0.005	-0.036
	(0.004)	(0.056)	(0.002)	(0.045)	(0.001)	(0.061)	(0.004)	(0.051)	(0.003)	(0.039)
African language school	-0.005	0.025	-0.013**	0.093	0.002	-0.004	-0.021***	0.083	-0.008	0.026
	(0.003)	(0.120)	(0.005)	(0.048)	(0.003)	(0.038)	(0.007)	(0.053)	(0.019)	(0.017)
School has a library	-0.001	0.015	-0.000	0.001	0.001	-0.021	0.002	0.048	0.000	0.064
	(0.001)	(0.025)	(0.001)	(0.010)	(0.003)	(0.024)	(0.004)	(0.030)	(0.002)	(0.036)
School has computers	0.003	-0.037	-0.001	-0.016	-0.001	0.002	0.000	0.028	0.004	0.063
	(0.004)	(0.024)	(0.002)	(0.012)	(0.002)	(0.013)	(0.001)	(0.033)	(0.005)	(0.031)
Constant		-0.283		-0.247		-0.492***		-0.359***		-0.579***
		(0.158)		(0.114)		(0.155)		(0.107)		(0.118)
Total	-0.135***	-0.368***	-0.161***	-0.224***	-0.196***	-0.296***	-0.200***	-0.356***	-0.168***	-0.230***
	(0.025)	(0.043)	(0.014)	(0.029)	(0.022)	(0.027)	(0.022)	(0.032)	(0.031)	(0.031)
N	2,024		2,256		2,560		2,503		2,391	

Notes: All models control for province. Standard errors are calculated at the school level and reported in parentheses. *** p < 0.01; ** p < 0.05; * p < 0.1.

Table C15: Results from the first decomposition (includes overage students): TIMSS Grade 5 (Mathematics)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5	
Gender gap	-0.220*** (0.058)		-0.215*** (0.030)		-0.211*** (0.024)		-0.149*** (0.027)		-0.002 (0.071)	
Boys' average	-0.618*** (0.079)		-0.520*** (0.040)		-0.288*** (0.028)		0.168*** (0.042)		0.968*** (0.080)	
Girls' average	-0.398*** (0.053)		-0.306*** (0.042)		-0.077*** (0.030)		0.282*** (0.043)		0.970*** (0.098)	
	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained
Overage	-0.014 (0.008)	-0.033 (0.030)	-0.052*** (0.007)	0.011 (0.017)	-0.030*** (0.005)	0.023 (0.016)	-0.038*** (0.006)	-0.009 (0.013)	-0.052*** (0.014)	0.020 (0.017)
Confidence index	-0.006 (0.007)	-0.019 (0.013)	-0.009 (0.005)	0.005 (0.005)	-0.006 (0.004)	0.000 (0.002)	-0.043*** (0.009)	0.000 (0.002)	0.023 (0.013)	0.017 (0.012)
Engagement index	0.002 (0.006)	-0.009 (0.009)	-0.007*** (0.002)	0.017** (0.008)	-0.004 (0.002)	0.002 (0.002)	-0.019*** (0.007)	-0.006 (0.006)	-0.000 (0.000)	-0.038*** (0.011)
Enjoyment index	-0.014 (0.009)	0.024 (0.012)	-0.022** (0.008)	0.001 (0.009)	-0.010 (0.006)	0.000 (0.001)	-0.036** (0.006)	0.010 (0.006)	0.003 (0.006)	-0.001 (0.009)
Belonging index	-0.009 (0.005)	0.001 (0.002)	-0.002 (0.003)	-0.004 (0.004)	-0.006** (0.003)	-0.002 (0.002)	-0.006 (0.003)	-0.004 (0.004)	-0.003 (0.004)	0.015 (0.008)
Bullying index	-0.005 (0.005)	-0.008 (0.006)	-0.021*** (0.006)	-0.011 (0.005)	-0.014*** (0.004)	-0.003 (0.002)	-0.016*** (0.005)	-0.003 (0.003)	-0.015** (0.006)	-0.013 (0.014)
Homework	-0.002 (0.003)	-0.118** (0.042)	-0.004 (0.003)	0.030 (0.028)	-0.009** (0.004)	0.012 (0.029)	-0.018*** (0.005)	0.041 (0.038)	-0.012** (0.006)	0.089 (0.058)
Attended ECD	-0.002 (0.003)	0.015 (0.025)	-0.004** (0.002)	-0.046 (0.040)	-0.007*** (0.002)	0.065** (0.025)	-0.000 (0.000)	0.025 (0.035)	-0.014*** (0.004)	-0.016 (0.040)
Asset index	-0.011 (0.007)	0.015 (0.025)	0.000 (0.001)	-0.010*** (0.003)	-0.000 (0.001)	-0.000 (0.001)	0.007** (0.003)	-0.012 (0.018)	0.001 (0.002)	-0.051 (0.042)
First language	-0.005 (0.004)	0.022 (0.018)	-0.004 (0.003)	-0.036*** (0.010)	-0.000 (0.002)	0.009 (0.018)	-0.009** (0.004)	-0.017 (0.026)	-0.034** (0.015)	-0.071 (0.038)
School has a library	0.003 (0.003)	0.007 (0.013)	-0.000 (0.001)	0.020 (0.013)	-0.030*** (0.005)	0.038** (0.018)	0.002 (0.004)	-0.006 (0.021)	-0.052*** (0.014)	0.067 (0.116)
School has computers	0.001 (0.003)	0.029 (0.015)	-0.001 (0.002)	-0.024 (0.011)	-0.006 (0.004)	0.009 (0.022)	0.000 (0.001)	-0.052** (0.020)	0.023 (0.013)	0.025 (0.092)
Constant		-0.016 (0.109)		-0.042 (0.069)		-0.305*** (0.033)		-0.033 (0.058)		0.027 (0.167)
Total	-0.044** (0.018)	-0.172** (0.065)	-0.120*** (0.020)	-0.088*** (0.025)	-0.072*** (0.014)	-0.151*** (0.027)	-0.089*** (0.022)	-0.066*** (0.014)	-0.087 (0.053)	0.077 (0.054)
N	1,802		2,084		2,207		2,280		1,774	

Notes: All models control for province. Standard errors are calculated at the school level and reported in parentheses. *** p < 0.01; ** p < 0.05; * p < 0.1.

Table C16: Selected results from model including additional school-level covariates (PIRLS)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Gender gap	-0.495***	-0.381***	-0.438***	-0.464***	-0.309***
	(0.055)	(0.032)	(0.033)	(0.036)	(0.046)
Explained	-0.093***	-0.112***	-0.144***	-0.141***	-0.093***
	(0.027)	(0.015)	(0.026)	(0.023)	(0.032)
Unexplained	-0.402***	-0.269***	-0.294***	-0.323***	-0.216***
	(0.043)	(0.027)	(0.025)	(0.035)	(0.033)
N	1,347	1,514	1,698	1,707	1,756

Notes: All models include the same set of controls included in the main estimation (Section 4.5.), as well as teacher's gender, age, and years of experience, whether the class has a library corner, and whether students can borrow books from the classroom library. Standard errors are calculated at the school level and reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table C17: Selected results from model including additional school-level covariates (TIMSS)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Gender gap	-0.176***	-0.182***	-0.211***	-0.114***	0.147
	(0.044)	(0.043)	(0.028)	(0.029)	(0.092)
Explained	-0.047**	-0.077***	-0.043	-0.035	0.085
	(0.022)	(0.019)	(0.020)	(0.022)	(0.049)
Unexplained	-0.129***	-0.106***	-0.168***	-0.079***	0.061
	(0.039)	(0.035)	(0.022)	(0.020)	(0.069)
N	1,354	1,644	1,577	1,756	1,478

Notes: All models include the same set of controls included in the main estimation (Section 4.5.), as well as teacher's gender, age, and years of experience. Standard errors are calculated at the school level and reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table C18: Decomposition results – Full samples, including parental education in asset index

	PIRLS (Reading) Grade 4					TIMMS (Mathematics) Grade 5				
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Gender gap	-0.553***	-0.541***	-0.409***	-0.546***	-0.454***	-0.178***	-0.281***	-0.209***	-0.168***	0.003
	(0.060)	(0.053)	(0.057)	(0.058)	(0.064)	(0.045)	(0.049)	(0.027)	(0.019)	(0.066)
Boys' average	-0.572***	-0.352***	-0.250***	-0.189***	0.268***	-0.656***	-0.554***	-0.265***	0.039	0.938***
	(0.059)	(0.052)	(0.060)	(0.066)	(0.101)	(0.069)	(0.041)	(0.026)	(0.030)	(0.080)
Girls' average	-0.019	0.189***	0.159***	0.357***	0.722***	-0.478***	-0.273***	-0.056	0.206***	0.935***
	(0.058)	(0.046)	(0.048)	(0.062)	(0.100)	(0.049)	(0.049)	(0.032)	(0.027)	(0.091)
	Explained	Explained	Explained	Explained	Explained	Explained	Explained	Explained	Explained	Explained
Overage	-0.036***	-0.036***	-0.038***	-0.038***	-0.032**	-0.021**	-0.041***	-0.036***	-0.041***	-0.050***
	(0.012)	(0.009)	(0.010)	(0.012)	(0.013)	(0.008)	(0.009)	(0.005)	(0.008)	(0.015)
Confidence index	-0.032***	-0.033***	-0.025	-0.067***	-0.065***	-0.002	-0.007	-0.012***	0.002	0.021
	(0.010)	(0.012)	(0.014)	(0.017)	(0.020)	(0.005)	(0.007)	(0.003)	(0.003)	(0.014)
Engagement index	-0.015	-0.012	-0.032**	-0.029***	-0.007	0.000	-0.009**	-0.004	-0.003	0.000
	(0.011)	(0.007)	(0.013)	(0.010)	(0.008)	(0.005)	(0.003)	(0.002)	(0.002)	(0.001)
Enjoyment index	-0.039***	-0.037***	-0.036**	-0.019**	-0.024**	-0.017	-0.027**	-0.006	-0.014**	0.006
	(0.012)	(0.011)	(0.015)	(0.009)	(0.010)	(0.009)	(0.011)	(0.006)	(0.005)	(0.005)
Belonging index	-0.007	-0.011	0.002	0.006	-0.002	-0.008	-0.004	-0.010**	-0.001	-0.004
	(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)	(0.004)	(0.001)	(0.004)
Bullying index	-0.008	-0.010	-0.027***	-0.011	-0.011	-0.002	-0.019**	-0.010***	-0.021***	-0.017***
	(0.006)	(0.005)	(0.009)	(0.006)	(0.006)	(0.004)	(0.008)	(0.002)	(0.004)	(0.006)
Homework	-0.023***	-0.018***	-0.033***	-0.021**	-0.023***	0.000	0.000	-0.005***	-0.008**	-0.012***
	(0.008)	(0.006)	(0.009)	(0.009)	(0.009)	(0.000)	(0.002)	(0.002)	(0.003)	(0.004)
Attended ECD	-0.003	0.000	-0.000	-0.002	0.001	-0.002	-0.000	0.000	0.000	0.002
	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.000)	(0.000)	(0.000)	(0.003)
Asset index	-0.007	0.001	-0.001	0.005	0.002	0.007	0.004	0.001	-0.006	0.006
	(0.005)	(0.002)	(0.004)	(0.005)	(0.004)	(0.008)	(0.003)	(0.002)	(0.003)	(0.008)
First language	-0.001	-0.001	0.000	-0.000	-0.000	-0.009	-0.000	-0.001	-0.004	-0.026**
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.005)	(0.003)	(0.002)	(0.007)	(0.013)
Constant	-0.379	-0.119	-0.392	-0.627**	-0.071	-0.078	-0.193	-0.162***	0.038	0.085
	(0.345)	(0.278)	(0.218)	(0.276)*	(0.273)	(0.080)	(0.144)	(0.046)	(0.036)	(0.139)
Total	-0.177***	-0.161***	-0.189***	-0.191***	-0.212***	-0.050**	-0.103***	-0.084***	-0.094***	-0.061
	(0.025)	(0.028)	(0.034)	(0.040)	(0.046)	(0.020)	(0.024)	(0.012)	(0.014)	(0.049)
N	2,562	2,564	2,561	2,561	2,562	2,213	2,156	2,149	2,202	2,166

Notes: All models include controls for the language of the test, whether the school has a library and computers, and province. Standard errors are calculated at the school level and reported in parentheses. *** p < 0.01; ** p < 0.05; * p < 0.1.

Table C19: Decomposition results – Full samples, missing values on student attitudes zero-imputed

	PIRLS (Reading) Grade 4					TIMMS (Mathematics) Grade 5				
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Gender gap	-0.433***	-0.406***	-0.514***	-0.528***	-0.423***	-0.186***	-0.203***	-0.197***	-0.190***	-0.056***
	(0.045)	(0.037)	(0.036)	(0.048)	(0.046)	(0.034)	(0.041)	(0.036)	(0.033)	(0.071)
Boys' average	-0.481***	-0.387***	-0.281***	-0.321***	0.364***	-0.646***	-0.444***	-0.216***	0.024	0.891***
	(0.064)	(0.069)	(0.057)	(0.056)	(0.090)	(0.057)	(0.066)	(0.060)	(0.060)	(0.094)
Girls' average	-0.048	0.019	0.233***	0.206***	0.787***	-0.460***	-0.241***	-0.019	0.214***	0.947***
	(0.051)	(0.053)	(0.056)	(0.053)	(0.079)	(0.048)	(0.060)	(0.047)	(0.061)	(0.087)
	Explained	Explained	Explained	Explained	Explained	Explained	Explained	Explained	Explained	Explained
Overage	-0.031***	-0.045***	-0.041***	-0.044***	-0.033***	-0.032***	-0.037***	-0.038***	-0.051***	-0.048***
	(0.008)	(0.010)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.010)	(0.013)
Confidence index	-0.048***	-0.021***	-0.043***	-0.038***	-0.076***	-0.006	-0.009	-0.014**	-0.005	0.014
	(0.012)	(0.008)	(0.010)	(0.010)	(0.016)	(0.005)	(0.007)	(0.006)	(0.006)	(0.014)
Engagement index	-0.014***	-0.029***	-0.013**	-0.021***	-0.019***	-0.005	-0.013**	-0.004	-0.002	-0.000
	(0.006)	(0.008)	(0.005)	(0.006)	(0.006)	(0.003)	(0.005)	(0.003)	(0.003)	(0.000)
Enjoyment index	-0.021***	-0.025***	-0.048***	-0.031***	0.012	-0.022***	-0.025**	-0.016***	-0.022***	0.002
	(0.007)	(0.008)	(0.011)	(0.008)	(0.010)	(0.007)	(0.010)	(0.008)	(0.009)	(0.005)
Belonging index	-0.005	-0.013***	-0.018***	-0.015***	0.009	-0.011**	-0.005	-0.009	-0.004	-0.000
	(0.004)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)
Bullying index	-0.003	-0.009	-0.015***	-0.010***	-0.016***	-0.008	-0.019***	-0.014**	-0.024***	-0.016***
	(0.003)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)
Homework	-0.015***	-0.024***	-0.013***	-0.015***	-0.056***	-0.001	-0.001	-0.006	-0.008**	-0.010**
	(0.005)	(0.007)	(0.005)	(0.005)	(0.014)	(0.001)	(0.002)	(0.004)	(0.004)	(0.005)
Attended ECD	-0.000	-0.000	-0.000	0.001	0.005**	-0.002	0.000	0.000	-0.001	-0.000
	(0.001)	(0.002)	(0.000)	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Asset index	0.003	0.003	-0.003	0.005	0.012**	0.011**	0.004	0.001	-0.002	0.002
	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)	(0.005)	(0.002)	(0.003)	(0.002)	(0.005)
First language	0.004	0.001	0.000	-0.007	-0.007	-0.005	0.001	-0.000	-0.002	-0.023
	(0.004)	(0.002)	(0.002)	(0.005)	(0.005)	(0.004)	(0.002)	(0.001)	(0.005)	(0.016)
Constant	-0.129***	-0.161***	-0.203***	-0.176***	-0.170***	0.128	0.063	-0.140	-0.512	-0.571
	(0.022)	(0.020)	(0.023)	(0.024)	(0.030)	(0.173)	(0.250)	(0.186)	(0.304)	(0.365)
Total	-0.129***	-0.161***	-0.203***	-0.176***	-0.170***	-0.077***	-0.104***	-0.100***	-0.125***	-0.101
	(0.022)	(0.020)	(0.023)	(0.024)	(0.030)	(0.021)	(0.023)	(0.020)	(0.022)	(0.058)
N	2,592	2,576	2,563	2,525	2,554	2,213	2,200	2,149	2,202	2,168

Notes: All models include controls for the language of the test, whether the school has a library and computers, and province. Standard errors are calculated at the school level and reported in parentheses. *** p < 0.01; ** p < 0.05; * p < 0.1